PTO 02-2534

CY=FR DATE=19980108 KIND=A1
PN=2,765,588

RETROVIRAL NUCLEAR MATERIAL AND NUCLEOTIDE FRAGMENTS ESPECIALLY ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PREVENTIVE AND THERAPEUTIC PURPOSES [MATERIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIES A LA SCLEROSE EN PLAQUES ET/OU LA POLYARTHRITE RHUMATOIDE, A DES FINS DE DIAGNOSTIC, PROPHYLACTIQUES ET THERAPEUTIQUES]

Bio Merieux S.A.

UNITED STATES PATENT AND TRADEMARK OFFICE
Washington, D.C. MAY 2002

Translated by: FLS, Inc.

PUBLICATION COUNTRY	(10):	France
DOCUMENT NUMBER	(11):	2,765,588
DOCUMENT KIND	(12):	A1
PUBLICATION DATE	(43):	19990108
PUBLICATION DATE	(45):	
APPLICATION NUMBER	(21):	97 08816
APPLICATION DATE	(22):	19970707
ADDITION TO	(61):	
INTERNATIONAL CLASSIFICATION	(51):	C 12 N 15/48
DOMESTIC CLASSIFICATION	(52):	
PRIORITY COUNTRY	(33):	FR
PRIORITY NUMBER	(31):	
PRIORITY DATE	(32):	
DESIGNATED CONTRACTING STATES	(84):	
INVENTORS	(72):	NA
APPLICANT	(71):	Bio Merieux S.A.
TITLE	(54):	RETROVIRAL NUCLEAR MATERIAL AND NUCLEOTIDE FRAGMENTS ESPECIALLY ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PREVENTIVE AND THERAPEUTIC PURPOSES
FOREIGN TITLE	(54A):	Materiel nucleique retroviral et fragments nucleotidiques notamment associes a la sclerose in plaques et/ou la polyarthrite rhumatoide, a des fins de diagnostic, prophylactiques et therapeutiques

Retroviral Nuclear Material and Nucleotide Fragments Especially Associated with Multiple Sclerosis and/or Rheumatoid Arthritis, for Diagnostic, Preventive and Therapeutic Purposes

Abstract

Nuclear material, in the isolated or purified state, and nucleotide fragment, which includes a nucleotide fragment chosen in the group that consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the complementary sequences of the sequences (i); and (iii) the sequences equivalent to the sequences (i) or (ii), in particular the sequences that have for an entire series 100 contiguous monomers, at least 50%, and preferably at least 70% homology with the sequences (i) or (ii) respectively, and their uses for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis.

Multiple sclerosis (MS) is a myelin destroying disease of the /1* central nervous system (CNS) whose full cause is still unknown.

Many studies have supported the hypothesis of a viral etiology of the disease, but none of the tested known viruses has turned out to be the sought causal agent: a review of the viruses studied over many years in MS was done by E. Norrby and R.T. Johnson.

Recently a retrovirus, different from the known human retroviruses, was isolated in patients afflicted with MS. The authors were also able to show that this retrovirus could be transmitted in vitro, that the MS afflicted patients produced antibodies capable of recognizing proteins associated with the infection of the leptomeningeal cells of this retrovirus, and that the expression of the latter could be strongly stimulated by the proximate precocious genes of certain herpes viruses.

All these results argue in favor of the role of at least one unknown retrovirus or of a virus that has reverse transcriptase (RT) activity detectable by the method published by H. Perron and classified as "LM7 type RT" activity in MS.

---

*Number in the margin indicates pagination in the foreign text.

The studies by the applicant allowed us to obtain two continuous lines of cells infected by natural isolated cultures coming from two different patients afflicted with MS, by a culture procedure, such as described in the document WO-A-93 20188, in which the content is incorporated by reference to the present description. These two lines, derived from cells of human choroidal plexus, named LM7PC and PLI-2, were deposited at the ECACC on July 22, 1992 and January 8, 1993 respectively, under the numbers 92 072201 and 93 010817, in conformity with the stipulations of the Budapest Treaty. In addition, the viral isolated cultures that have RT activity of the LM7 type were also deposited at the ECACC under the general designation of "strains." The "strain" or isolated culture harbored by the PLI-2 strain, named POL-2, was filed at the ECACC on July 22, 1992 under the number V92072202. The "strain" or isolated culture harbored by the line LM7PC, named MS7PG, was filed at the ECACC on January 8, 1993 under number V93010816. /2

From the aforementioned cultures and isolated material, characterized by some biological and morphological criteria, we set out to characterized the nuclear material associated with the viral particles produced in these cultures.

The portions of the genome already described were used to perfect molecular detection tests of the viral genome and some immuno-serological tests, using the amino acid sequences coded by the nucleotide sequences of the viral genome, in order to detect the immune response directed against epitopes associated with the infection and/or the viral expression.

These tools already let us confirm an association between MS and the expression of the sequences identified in the previously cited patents. However, the viral system discovered by the applicant is related to a complex retroviral system. Indeed, the sequences found encapsulated in the extra-cellular particles produced by the different cell cultures of patients afflicted with MS show clearly that there is co-encapsulation of retroviral genomes that are related but different from the "wild" retroviral genome that produces the infecting viral particles. This phenomenon has been observed among the replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous. The notion of endogenous retrovirus is very important in the context of our discovery because, in the case of MSRV-1, we have observed that some endogenous retroviral sequences that include sequences homologous with the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV), related to MSRV-1 by all or part of their genome, explains the fact that the expression of the retrovirus MSRV-1 in human /3

cells can interact with similar endogenous sequences. These sequences are found in the case of pathogenic endogenous retroviruses and/or infections ones (for example, some ecotropic strains of murine leukemia virus), in the case of exogenous retrovirus in which the nucleotide sequence can be found partially or completely in the form of ERV's, in the genome of the host animal (ex. exogenic virus of breast tumor of the mouse transmitted by milk). These interactions consist mainly in (i) a transactivation or co-activation of ERVs by the replicative retrovirus, (ii) an "illegitimate" encapsulation of RNA related to ERVs, or of ERFs -rather cell RNA- that have simply some compatible encapsulation sequences, in the retroviral particles produced by the expression of the replicative strain, sometimes transmissible and sometimes with a characteristic pathogenicity, and (iii) some more or less important recombinations among the co-encapsulated genomes, especially in the reverse transcription phases, which lead to the formation of hybrid genomes, sometimes transmissible and sometimes with characteristic pathogenicity.

Thus, (i) different sequences related to MSRV-1 have been found in the purified viral particles; (ii) a molecular analysis of the different regions of the MSRV-1 retroviral genome must be done by analyzing systematically the co-encapsulated, interfering and/or recombined sequences that are generated by infection and/or expression of MSRV-1, and in addition, some clones can have parts of defective sequences produced by retroviral replication and matrix and/or transcription errors of reverse transcriptase; (iii) the families of sequences related to a single retroviral genomic region are the supports of global diagnostic detection that can be /4 optimized by the identification of unvarying regions among the clones expressed and by the identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides that can be produced only by a part, or even only one, of the clones expressed and under these conditions, the systematic analysis of the clones expressed in a region of given gene lets one evaluate the frequency of variation and/or recombination of the genome MSRV-1 in this region and to define the optimal sequences for applications, especially diagnostic; (iv) the pathology caused by a retrovirus such as MSRV-1 can be a direct effect of its expression and of the proteins or peptides produced due to this fact, but also an effect of the activation, the encapsulation, and the recombination of related or heterologous genomes and of proteins or peptides produced by these events; thus, these genomes associated with the expression of and/or the infection by MSRV-1 are an integral part of the potential pathogenicity of this virus and therefore, comprise supports for diagnostic detection and particular therapeutic targets. Also, any agent associated with, or a co-factor of these interactions responsible for the

pathogenicity in question, such as MSRV-2 or the gliotoxin factor described in the patent application published under number FR-2,716,198, can participate in the development of an overall and very effective strategy of diagnosis, prognosis, therapeutic monitoring and/or integrated therapy for MS especially, but also for any other disease associated with the same agents.

In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which was described in the French patent application filed under number 95 02960. This discovery shows that, by applying methodological approaches similar to those that were used in the studies of the applicant for MS, it /5 was possible to identify a retrovirus expressed in RA that shares the sequences described for MSRV-1 in MS and also the co-existence of an MSRV-2 associated sequence also described in MS. With respect to MSRV-1, the sequences jointly detected in MS and RA pertain to the genes pol and gag. Given the present state of knowledge, one can associate the described sequences gag and pol with the MSRV-1 strains expressed in these two diseases.

The present patent application has as one goal different results, which are supplementary with respect to those already protected by the French patent applications:

- No. 92 04322 of April 3, 1992, published under No. 2,689,519;
- No. 92 13447 of November 3, 1992, published under No. 2,689,521;
- No. 92 13443 of November 3, 1992, published under No. 2,689,520;
- No. 94 01529 of February 4, 1994, published under No. 2,715,936;
- No. 94 01531 of February 4, 1994, published under No. 2,715,939;
- No. 94 01530 of February 4, 1994, published under No. 2,715,936;
- No. 94 01532 of February 4, 1994, published under No. 2,715,937;
- No. 94 14322 of November 24, 1994, published under No. 2,727,428;
- No. 94 15810 of December 23, 1994, published under No. 2,728,585;

And

- The patent application WO-97/06260.

The present invention pertains first to a nuclear material that can consist of a retroviral material, in the isolated or purified state, which can be comprehended or characterized in different ways:

- It includes a nucleotide sequence chosen in the group that consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences that are complementary to the sequences of (i); and (iii) the sequences /6 equivalent to the sequences of (i) or (ii), in particular the sequences that have for the entire series of 100 contiguous monomers, at least 50%, and preferably at least 70% homology with

the sequences (i) or (ii) respectively;

- It codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with a peptide sequence chosen in the group that consists of SEQ ID NO: 113, SEQ ID NO 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- Its gene pol includes a nucleotide sequence that is identical or equivalent to a sequence chosen in the group that consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary sequences;

- The end 5″ of its gene pol begins at the nucleotide 1419 of SEQ ID NO: 130;

- Its gene pol codes for a polypeptide that has, for its entire continuous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with the peptide sequence SEQ ID NO: 113;

- The end 3′ of its gene gag ends at the nucleotide 1418 of SEQ ID NO: 130;

- Its gene env includes a nucleotide sequence identical to or equivalent to a sequence chosen in the group that consists of SEQ ID NO: 117, and its complementary sequences;

- Its gene env includes a nucleotide sequence that begins at the nucleotide 1 of SEQ ID NO: 117 and ends at the nucleotide 233 of SEQ ID NO: 114;

- Its gene env codes for a polypeptide that has, for all it contiguous series of at least 30 amino acids, at least 505, and preferably at least 70% homology, with the sequence SEQ ID NO: 118;

- The region U3R of its LTR 3′ includes a nucleotide sequence that terminates at the nucleotide 617 of SEQ ID NO: 114;

- The region RU5 of its LTR 5′ includes a nucleotide /7
sequence that begins at the nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NOT: 141 or SEQ ID NO: 142;

- A retroviral nuclear material that includes a sequence that begins at nucleotide 755 of SEQ ID NO: 120 and that terminates at nucleotide 617 of SEQ ID NO: 114;

- The retroviral nuclear material as defined previously is in particular associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

The invention pertains also to a nucleotide fragment that meets at least one of the following definitions:

- It includes or consists of a nucleotide sequence chosen in the group that consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the complementary sequences of the sequences (i); and (iii) the sequences equivalent to the sequences of (i) or (ii), in particular the sequences than have for the entire series of 100 contiguous monomers, at least 505, and preferably at least 70% homology with the sequences (i) or (ii) respectively;

- It includes or consists of a nucleotide sequence that codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with a peptide sequence chosen in the group that consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

Other aims of the present invention are the following:

- A nuclear probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, capable of hybridizing specifically on any fragment previously defined and belonging to the genome of the said retrovirus; it has advantageously from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides; /8

- A beginning for amplification by polymerization of RNA or DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, which includes a nucleotide sequence identical or equivalent to at least one part of the nucleotide sequence of a fragment defined previously, especially a nucleotide sequence that has for the entire series of 10 contiguous monomers, at least 50%, and preferably at least 70% homology with at least the said part of the said fragment; preferably the nucleotide sequence of a beginning of the invention is chosen among SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and SEQ ID NO: 133;

- A RNA or a DNA, and especially a replication and/or expression vector, which includes a genomic fragment of the nuclear

material or a fragment defined previously;

- A peptide coded by any open reading frame belonging to a previously defined nucleotide fragment, especially a polypeptide, an oligo-peptide for example that forms or includes an antigen determinant recognized by the sera of patients infected by the MSRV-1 virus, and/or in which the MSRV-1 virus has been reactivated; a preferred peptide includes a sequence that is identical, partially or fully, or equivalent to a sequence chosen among SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- A diagnostic, preventive, or therapeutic compound /9 especially for inhibiting the expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, which includes a previously defined nucleotide fragment;

- A procedure for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, which includes the stages that consist of putting RNA and/or DNA presumed to belong or to come from the said retrovirus in contact, or their RNA and/or complementary DNA, with a compound that includes a nucleotide fragment as defined earlier.

Before detailing the invention different terms used in the description and the claims will now be defined:

- By strain or isolated culture or isolated material we mean any infecting and/or pathogenic biological fraction, which contains viruses and/or bacteria and/or parasites for example, which generate a pathogenic and/or antigenic power, harbored by a culture or a living host; as an example, a viral strain according to the preceding definition can contain a co-infecting agent, a pathogenic unicellular organism;

- The term "MSRV" used in the present description designates any pathogenic agent and/or infecting agent, associated with MS, especially a viral species, the attenuated strains of the said viral species, or the interfering defective particles that contain encapsulated genomes or even some genomes recombined with one part of the MSRV-1 genome, derived from this species. It is known that the viruses and particularly the viruses that contain RNA have a variability, consecutive especially with some relatively high rates of spontaneous mutation, which will be considered subsequently for defining the concept of equivalence,

- By human virus we mean a virus capable of infecting or of being harbored by human beings,

- Considering all the variations and/or natural or induced recombinations, which could be met in practice of the present invention, the aims of the latter, defined previously and in the claims, have been expressed by including the equivalents defined subsequently, especially nucleotide or peptide homologous sequences, /10

- The variant of a virus or a pathogenic and/or infecting agent according to the invention includes at least one antigen recognized by at least one antibody directed against at least one correspondent antigen of the said virus and/or the said pathogenic and/or infecting agent, and/or a genome in which every part is detected by at least one hybridization probe, and/or at least one specific nucleotide amplification beginning of the said virus and/or pathogenic and/or infecting agent, under specific hybridization conditions well known by a man of the art,

- According to the invention, a nucleotide fragment or an oligo-nucleotide or a polynucleotide is in a series of monomers, or a biopolymer, characterized by the informational sequence of natural nucleic acids, capable of hybridizing with any other nucleotide fragment under predetermined conditions, the series capable of containing monomers with different chemical structures and ob being produced from one molecule of natural nucleic acid and/or by gene recombination and/or by chemical synthesis; a nucleotide fragment can be identical to a genome fragment of the MSRV-1 virus considered by the present invention, especially a gene of the latter, pol or env for example in the case of the said virus;

- Thus, a monomer can be a natural nucleotide of nucleic acid, in which the constituent elements are a sugar, a phosphate group and a nitrogenous base; in RNA the sugar is ribose, in DNA the sugar is desoxy-2-ribose; whether it is a question of DNA or RNA, the nitrogenous base is chosen among adenine, guanine, uracil, cytosine, thymine; or the nucleotide can be modified in at least one of the three constituent elements; as an example, the modification can occur at the level of the bases, generating modified bases such as inosine, methyl-5-desoxycytidine, desocyuridine, dimethyl amino-5-desoxyuridine, diamino-2,6-purine, bromo-5-desoxyuridine and any other modified base that promotes the hybridization; at the sugar level, the modification can consist in the replacement of at least one desoxyribose by a polyamide, and at the level of the phosphate group, the modification can consist of its replacement by some esters, especially chosen among the esters of diphosphate, alkyl, and arylphosphonate and phosphorothioate, /11

- By "informational sequence" we mean any ordered series of monomers whose chemical nature and order in a reference direction comprise or not functional information of the same quality as that of natural nucleic acids,

- By hybridization we mean the process during which, under appropriate operating conditions, two nucleotide fragments that have sufficiently complementary sequences pair to form a complex structure, especially a double or triple one, preferably in the form of a helix,

- A probe includes a nucleotide fragment synthesized by chemical means or obtained by digestion or enzymatic cutting of a longer nucleotide fragment, which includes at least six monomers, advantageously from 10 to 100 monomers, preferably from 10 to 30 monomers, and having hybridization specificity under specific conditions; preferably, a probe that has less than 10 monomers is not used alone, but is used in the presence of other probes with /12
size just as short or not; under certain particular conditions it could be useful to use probes of size greater than 100 monomers; a probe can in particular be used for diagnostic purposes and in this case one will use capture and/or detection probes, for example,

- The capture probe can be immobilized on a solid support by any suitable means, that is directly or indirectly, by covalence or passive adsorption for example,

- The detection probe can be marked by means of a marker chose especially among radioactive isotopes, enzymes especially chosen among peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorigenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorigenic or luminescent compounds, analogues of nucleotide bases, and biotin,

- The probes used for diagnostic purposes of the invention can be put to work in all the known hybridization techniques, and especially the techniques called "DOT-BLOT," "SOUTHERN BLOT," "NORTHERN BLOT," which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as the target, the SANDWICH technique; advantageously one uses the SANDWITCH technique in the present invention, which includes a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence at least partially different,

- Any probe according to the present invention can be hybridized in vivo or in vitro on RNA and/or on DNA, to block the replication phenomena, especially translation and/or transcription,

and/or to degrade the said DNA and/or RNA,

- An initiator is a probe that includes at least six /13
monomers, and advantageously from 10 to 30 monomers, which have a hybridization specificity under specific conditions, for the initiation of an enzymatic polymerization, in an amplification technique for example such as PCR (polymerase chain reaction), in an elongation process, such as sequencing, in a method of reverse transcription or similar method,

- Two nucleotide or peptide sequences are called equivalent or derived with respect to one another, or with respect to a reference sequence, if functionally the corresponding biopolymers can play approximately the same role, without being identical, vis-à-vis the application or use in question, or in the technique in which they occur; two sequences obtained due to the natural variability, especially spontaneous mutation of the species from which they have been identified, or induced, as well as two homologous sequences, the homology being defined subsequently, are equivalent in particular,

- By "variability" we mean any modification, spontaneous or induced by a sequence, especially by substitution, and/or insertion, and/or deletion of nucleotides and/or nucleotide fragments, and/or extension and/or shortening of the sequence at least at one of the ends; a non-natural variability can result from gene engineering techniques used, from the choice of synthesis initiators for example, degenerated or not, retained to amplify a nucleic acid; this variability can be conveyed by modifications of any initial sequence, considered as the reference, and capable of being expressed by a degree of homology with respect to the said reference sequence,

- The homology characterizes the degree of identity of two nucleotide or peptide fragments that are compared; it is measured by the identity percentage that is especially determined by direct /14
comparison of nucleotide or peptide sequences, with respect to reference nucleotide or peptide sequences,

- Any nucleotide fragment is called equivalent or derived from a reference fragment if it has a nucleotide sequence equivalent to the sequence of the reference fragment; based on the preceding definition the following in particular are equivalent to a reference nucleotide fragment:

(a) Any fragment capable of hybridizing at least partially with the complement of the reference fragment,

(b) Any fragment in which the alignment with the reference fragment results in revealing identical contiguous bases, with number greater that with any other fragment coming from another taxonomic group,

(c) Any fragment that results from or could result from the natural variability of the species, from which it is obtained,

(d) Any fragment that could result from gene engineering techniques applied to the reference fragment,

(e) Any fragment that includes at least eight continuous nucleotides that code for a peptide that is homologous or identical to the peptide coded by the reference fragment,

(f) Any fragment different from the reference fragment, by insertion, deletion, substitution of at least one monomer, extension, or shortening at least at one of its ends; for example, any fragment corresponding to the reference fragment, flanked at least at one of its ends by a nucleotide sequence that does not code for a polypeptide,

- By polypeptide we mean especially any peptide with at least two amino acids, especially an oligopeptide, protein, extract, separated, or substantially isolated or synthesized, by human intervention, especially those obtained by chemical synthesis, or by expression into a recombinant organism, /15

- By polypeptide coded in a partial manner by a nucleotide fragment we mean a polypeptide that has at least three amino acids coded by at least nine contiguous monomers included in the said nucleotide fragment,

- An amino acid is said to be an analogue of another amino acid when their respective physico-chemical characteristics such as polarity, hydrophobicity, and/or basicity, and/or acidity, and/or neutrality, are approximately the same; thus, a leucine is analogous to an isoleucine,

- Any polypeptide is said to be equivalent to or derived from a reference polypeptide if the compared polypeptides have approximately the same properties, and especially the same antigenic, immunological, enzymological properties, and the property of molecular recognition; it is especially equivalent to a reference polypeptide:

(a) Any polypeptide that has a sequence in which at least one amino acid has been substituted by an analogous amino acid,

(b) Any polypeptide that has an equivalent peptide sequence, obtained by natural or induced variation of the said reference polypeptide, and/or the nucleotide fragment that codes for the said polypeptide,

(c) A mimotope of the said reference polypeptide,

(d) Any polypeptide in the sequence of which one or several amino acids of the series L are replaced by an amino acid of the series D, and vice versa,

(e) Any polypeptide in the sequence of which a modifications of the lateral chains of the amino acids has been introduced, such as for example an acetylation of the amine functions, a carboxylation of the thiol functions, an esterification of the carboxylic functions, /16

(f) Any polypeptide in the sequence of which one of the peptide bonds has been modified, as for example the carba, retro, inverso, retro-inverso, reduced, and methylene-oxy bonds,

(g) Any polypeptide in which at least one antigen is recognized by antibodies directed against a reference polypeptide,

- The percentage of identity that characterizes the homology of two compared peptide fragments is at least 505 and preferably at least 70% according to the present invention.

Since a virus that has reverse transcriptase enzymatic activity can be genetically characterized just as easily in the RNA for as the DNA form, we should also make mention of DNA as well as viral RNA to characterize the sequences relative to a virus that has such reverse transcriptase activity, called MSRV-1 according to the present description.

The expressions of order used in the present description and the claims, such as "first nucleotide sequence" are not retained to express a particular order, but to define more clearly the invention.

By detection of a substance or agent we mean hereafter also an identification , a quantification, or a separation or isolation of the said substance or the said agent.

The invention will be better understood from reading the detailed description that follows in reference to the attached figures in which:

Figure 1 shows the general structure of proviral DNA and genomic RNA of MSRV-1.

Figure 2 shows the nucleotide sequence of the clone named CL6-5′ (SEQ ID No: 112) and three potential amino acid reading frames that are included under the nucleotide sequence. /17

Figure 3 shows the nucleotide sequence of the clone named CL6-3′ (SEQ ID No: 114) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 4 shows the nucleotide sequence of the clone called C15 (SEQ ID No: 117) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 5 shows the nucleotide sequence of the clone named 5M6 (SEQ ID No: 120) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 6 shows the nucleotide sequence of the clone named CL2 (SEQ ID No: 130) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 7 shows three potential amino acid reading frames expressed by pET28C-clone 2 and that are included under the nucleotide sequence.

Figure 8 shows three potential amino acid reading frames expressed by pER21C-clone 2 and that are included under the nucleotide sequence.

Figure 9 shows the nucleotide sequence of the clone named LB13 (SEQ ID No: 141) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 10 shows the nucleotide sequence of the clone named LA 15 (SEQ ID No: 142) and three potential amino acid reading frames that are included under the nucleotide sequence.

Figure 11 shows the nucleotide sequence of the clone named LB16 (SEQ ID No: 124) and three potential amino acid reading frames that are included under the nucleotide sequence.

EXAMPLE 1: PRODUCTION OF CL6-5′ REGION THAT CODES FOR THE - TERMINAL END OF THE INTEGRASE AND OF A REGION CL6-3′ THAT CONTAINS THE TERMINAL SEQUENCE 3′ OF THE GENOME MSRV-1 /18

A 3'RACE strain was carried out on total RNA extracted from plasma of a patient afflicted with MS. A healthy control plasma, treated under the same conditions, was used as a negative control. The synthesis of cDNA was accomplished with a dT oligo primer identified by SEQ ID No: 68 (5' GAC TCG CTG CAG ATC GAT TTT TTT TTT TTT TTT T 3')and the reverse transcriptase "Expand® RT" of Boehringer according to the conditions recommended by the company. A polymerase chain reaction (PCR) was carried out with the enzyme Klentaq (Clontech) under the following conditions: 94 degrees C for 5 minutes then 93 degrees for 1 minute, 58 degrees for 1 minute, 68 degrees for 3 minutes during 40 cycles and 68 degrees for 8 minutes, with a final reaction volume of 50 μl.

Primers used for the PCR:

- Primer 5', identified by SEQ ID No: 69 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3':

- Primer 3', identified by SEQ ID No: 68.

A second PCR called "semi-strain" was carried out with a primer 5' located inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 10 μl of amplification product derived from the first PCR.

Primers used for the semi-strain PCR:

- Primer 5', identified by SEQ ID No: 70

- 5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3''

- Primer 3', identified by SEQ ID No: 68.

The primers SEQ ID No: 69 and SEQ ID No: 70 are specific for the region pol of MSRV-1.

An amplification product of 1.9 Kb was obtained for the plasma of the MS patient. The corresponding fragment was not observed for the healthy control plasma. This amplification product was cloned /19 in the following way:

The amplified DNA was inserted into a plasmid by means of the TA Cloning® kit. The 2 μl of DNA solution were mixed with 5 μl of sterile distilled water, 1 microliter of a 10-fold concentrated bonding buffer solution "10X Ligation Buffer", 2 μl of "pCR® VECTOR (25 ng/ml) and 1 microliter of "T4 DNA LIGASE." This mixture was incubated overnight at 12 degrees C. The following stages were

carried out in conformity with the instruction of the TA Cloning® kit (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured in order to be grown and allow the extraction of the incorporated plasmids according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was interrupted by a suitable restriction enzyme and analyzed on agar gel. The plasmids that have an insert detected under ultraviolet light after marking of the gel with ethidium bromide were selected for the sequencing of the insert, following hybridization with a complementary primer of the promoter Sp6 present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for the use of the sequencing kit "PRISM® Ready Reaction AmpliTaq® FS, DyeDeoxy® Terminator" (Applied Biosystems, ref. 402119) and the automatic sequencing was accomplished on the devices 373 A and 377 Applied Biosystems, according to the instructions of the manufacturer.

The resulting clone contains a region CL6-5′ that codes for the N terminal end of the integrase and a region CL6-3′, which corresponds to the terminal region 3′ of MSRV-1 and which allows one to define the end of the envelope (234 pb) and the regions U3, R (401 pb) of the retrovirus MSRV1.

The region corresponding to the N terminal end of the integrase is represented by its nucleotide sequence (SEQ ID No: 112) in Fig. 1. The three potential reading frames are presented by their amino acid sequence under the nucleotide sequence, and the amino acid sequence of the N terminal end of the integrase is identified by SEQ ID No: 113. /20

The region C16-3′ is represented by its nucleotide sequence (SEQ ID No: 114) in Fig. 3. The three potential reading frames are presented by their amino acid sequence under the nucleotide sequence. An amino acid sequence that corresponds to the C-terminal end of the protein env of MSRV-1 is identified by SEQ ID No: 115.

EXAMPLE 2: PRODUCTION OF THE CLONE C15 THAT CONTAINS THE REGION THAT CODES FOR ONE PART OF THE ENVELOPE OF THE RETROVIRUS MSRV-1

A RT-PCR was carried out on the total RNA extracted from virions concentrated by ultra-centrifuging from the surface fluid of a culture of synoviocytes coming from a RA patient. The synthesis of cDNA was carried out with a primer dT oligo and the reverse transcriptase "Expand® RT" of Boehringer according to the conditions recommended by the company. A PCR was carried out with

the Expand® Long Template PCR System (Boehringer) under the following conditions: 94 degrees C for 5 minutes then 93 degrees for 1 minute, 60 degrees C for 1 minute, 68 degrees C for 3 minutes during 40 cycles and 68 degrees C for 8 minutes and with a final reaction volume for the PCR:

- Primer 5', identified by SEQ ID No: 69
- 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- Primer 3', identified by SEQ ID NO: 116
- 5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

A second PCR called "semi-strain" was carried out with a primer 5' located inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR (except that 30 cycles were carried out in place of 40), using 10 μl of the amplification product derived from the first PCR. /21

Primers used for the semi-strain PCR:

- Primer 5', identified by SEQ ID No: 70
- 5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';
- Primer 3', identified by SEQ ID No: 116

The primers SEQ ID No: 69 and SEQ ID No: 70 are specific for the region pol of MSRV-1. The primer SEQ ID NO: 116 is specific for the sequence FBd13 (also named B13) and is localized in the env region preserved among the onco-retroviruses.

An amplification product of 1932 pb was obtained and cloned in the following way: The amplified DNA was inserted in a plasmid by the help of the TA Cloning® kit. The different stages were carried out in conformity with the instructions of the TA Cloning® kit (Invitrogen). At the end of the procedure the white colonies of recombinant bacteria (white) were subcultured to be grown and to allow the extraction of the incorporated plasmids according to the so-called "miniprep" procedure. The preparation of plasmid of each recombinant colony was interrupted by a suitable restriction enzyme and analyzed on agar gel. The plasmids that have an insert detected under ultra-violet light after marking of the gel in ethidium bromide, were selected for the sequencing of the insert, following hybridization with a complementary primer of the promoter SP6 present on the cloning plasmid of the TA cloning kit®. The

reaction prior to the sequencing was then carried out according to the method recommended for the use of the "PRISM® Ready Reaction AmpliTaq® FS, DyeDeoxy® Terminator" (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out one the devices 373 A and 377 Applied Biosystems, according to the instructions of the manufacturer.

The clone C15 obtained contains a region that corresponds to /22 the region of the envelope of MSRV-1,with 1481 pb.

The region env of the clone C15 is represented by its nucleotide sequence (SEQ ID No: 117) in Fig. 5. The three potential reading frames of this clone are presented by their amino acid sequence under the nucleotide sequence. The reading frame corresponding to a structural env protein of MSRV-1 is identified by SEQ ID No: 118.

EXAMPLE 3: PRODUCTION OF A CLONE 5M6 THAT CONTAINS THE SEQUENCES OF THE TERMINAL 3' REGION OF THE ENVELOPE, FOLLOWED BY THE SEQUENCES U3, R, U5 OF THE MSRV-1 PROVIRAL TYPE.

A single-direction PCR was carried out on DNA extracted from B-lymphocytes immortalized in the culture of a RA patient. The PCR was carried out with Expand® Long Template PCR System (Boehringer) under the following conditions: 94 °C for 3 minutes then 93 °C for 1 minute for 10 cycles, then 93 °C for 1 minute, 60 °C for 1 minute with 15 seconds extension for each cycle, 68 degrees C for 3 minutes for 35 cycles and 68 °C for 7 minutes and with a final reaction volume of 50 μl.

The primer used for the PCR identified by SEQ ID No: 119 is 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT AAC CG 3';

The primers SEQ ID NO: 119 is specific for the region env of the clone C15.

An amplification product of 1673 pb was obtained and cloned in the following way: The amplified DNA was inserted in a plasmid by the help of the TA Cloning® kit. The different stages were carried out in conformity with the instruction of the TA Cloning® kit (Invitrogen). At the end of the procedure the white colonies of recombinant bacteria (white) were cultured again in order to be cultivated and to allow the extraction of the plasmids /23 incorporated according to the so-called "miniprep" procedure. The preparation of plasmid of each recombinant colony was cut by a suitable restriction enzyme and analyzed on agar gel. The plasmids that have an insert detected under ultra-violet light after marking of the gel with ethidium bromide were selected for the sequencing

of the insert, after hybridization with a complementary primer of the promoter T7 present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for the use of the "PRISM® sequencing kit Ready Reaction AmpliTaq® FS, DyeDeoxy® Terminator" (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the devices 373 a and 377 Applied Biosystems, according to the instructions of the manufacturer.

The resulting clone 5M6 contains a region that corresponds to the region 3' of the envelope of MSRV-1, with 492 pb followed by regions U3, R and U5 (837 pb) of MSRV1.

The clone 5M6 is represented by its nucleotide sequence (SEQ ID No: 120) in Fig. 7. The three potential reading frames of this clone are presented by their amino acid sequence under the nucleotide sequence. The reading frame corresponding to the C-terminal end of the protein env MSRV-1 is identified by SEQ ID No: 121.

EXAMPLE 4: PRODUCTION OF THE CLONE LB16 THAT CONTAINS THE REGION THAT CODES THE INTEGRASE OF THE MSRV-1 RETROVIRUS.

An RT-PCR was carried out on the total RNA treated with the DNAseI and extracted from a plexus choroideus that comes from an MS patient. The synthesis of cDNA was carried out with a dT oligo primer and the "Expand® RT" reverse transcriptase of Boehringer according to the conditions recommended by the company. A "no RT" control was carried out at the same tine on the same material. A /24 PCR was carried out with the Taw polymerase (Perkin Elmer) under the following conditions: 95 °C for 5 minutes then 95 °C for 1 minute, 55 °C for 1 minute, 72 °C for 2 minutes during 35 cycles and 72 °C for 8 minutes and with a final reaction volume of 50 μl.

Primers used for the PCR:

- Primer 5', identified by SEQ ID No: 122
- 5' GGC ATT GAT AGC ACC CAT CAG 3';
- Primer 3', identified by SEQ ID No: 123
- 5' CAT GTC ACC AGG GTG GAA TAG 3'

The primer SEQ ID No: 122 is specific for the region pol of MSRV-1 and more precisely similar to the integrase region described previously. The primer SEQ ID No 123 has been defined on some sequences of clones obtained during prior tests.

An amplification product of about 760 pb was obtained only in the test with RT and was cloned in the following way:

The amplified DNA was inserted in a plasmid by the help of the TA Cloning® kit. The different stages were carried out in conformity with the instructions of the TA Cloning® kit (Invitrogen). At the end of the procedure the white colonies of recombinant bacteria (white) were subcultured in order to be cultivated and to allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The preparation of plasmid of each recombinant colony was cut off by a suitable restriction enzyme and analyzed on agar gel. The plasmids that have an insert detected under ultra-violet light after marking of the gel in ethidium bromide were selected for the sequencing of the insert, following hybridization with a complementary primer of the promoter T7 present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for the use of the "PRISM® Ready Reaction AmpliTaq® FS DyeDeoxy® Terminator" sequencing kit (Applied Biosystems, ref. 402119) and the automatic sequencing was carried /25 out one the devices 373 A and 377 Applied Biosystems, according to the instructions of the manufacturer.

The clone LB16 produced contains the sequences corresponding to the integrase. The nucleotide sequence of this clone is identified by SEQ ID No: 124 in Fig. 11, three reading frames were determined.

EXAMPLE 5: PRODUCTION OF A CLONE 2, CL2, WHICH CONTAIN AT 3' A PART THAT IS HOMOLOGOUS TO THE GENE POL, WHICH CORRESPONDS TO THE PROTEASE GENE, AND TO THE GENE GAG (GM3) THAT CORRESPONDS TO THE NUCLEOCAPSIDE, AND A NEW CODING REGION 5' THAT CORRESPONDS TO THE GENE GAG MORE SPECIFICALLY THE MATRIX AND THE CAPSIDE OF MSRV-1.

An amplification by PCR was carried out on total RNA extracted from 100 μl of plasma of a patient afflicted with MS. A water control, treated under the same conditions, was used as a negative control. The synthesis of cDNA was carried out with 300 pmole of a random primer (GIBSO-BRL, France) and the reverse transcriptase "Expand RT" (Boehringer Mannheim, France) according to the conditions recommended by the company. An amplification by PCR was carried out with the enzyme Taw polymerase (Perkin Elmer, France) using 10 μl of cDNA under the following conditions: 94 °C for 2 minutes, 55 °C for 1 minute, 72 degrees for 2 minutes then 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 2 minutes during 30 cycles and 72 °C during 7 minutes and with a final reaction volume of 50 μl.

Primers used for the amplification by PCR:

- Primer 5', identified by SEQ ID No: 126
- 5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
- Primer 3', identified by SEQ ID NO: 127
- 5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

A second amplification by PCR called "semi-strain" was carried out with a primer 5' located inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 10 μl of the amplification product derived from the first PCR. /26

Primers used for the amplification by PCR semi-strain:

- Primer 5', identified by SEQ ID No: 128
- 5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3';
- Primer 3', identified by SEQ ID No: 129
- 5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

The primers SEQ ID No: and SEQ ID No: are specific for the region pol, clone G+E+A, more specifically the region E: nucleotide position No 423 to no. 448. The primers used in the region 5' were defined on some sequences of clones obtained during prior tests.

An amplification product of 1511 pb was obtained from the RNA extracted from the plasma of an MS patient. The corresponding fragment was not observed for the water control. This amplification product was cloned in the following way.

The amplified DNA was inserted in a plasmid by the help of the TA Cloning® kit. Two μl of the DNA solution were mixed with 5 μl of sterile distilled water, 1 μl of a binding buffer solution concentrated 10X "10X Ligation Buffer," 2 μl of "PCR® VECTOR" (25 ng/ml) and 1 microliter of "T4 DNA LIGASE." This mixture was incubated overnight at 14 °C. The following stages were carried out in conformity with the instructions of the TA Cloning® kit (Invitrogen). The mixture was spread out after transformation of the ligation in some E. coli INVαF' bacteria. At the end of the procedure the white colonies of recombinant bacteria were subcultured in order to be cultivated and to allow the extraction of the plasmids incorporated according to the so-called /27

"minipreparation of DNA" procedure (17). The plasmid preparation of each recombinant colony was cut by the restriction enzyme Eco RI and analyzed on agar gel. The plasmids that have an insert detected under ultra-violet light after marking of the gel with ethidium bromide were selected for the sequencing of the insert, after hybridization with a complementary primer of the promoter T7 present on the cloning plasmid of the TA Cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for the use of the "PRISM® Ready Reaction Amplitaq®FS, DeyDeoxy® Terminator" sequencing kit (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the devices 373 A and 377 Applied Biosystems, according to the instructions of the manufacturer.

The resulting clone, named CL2, contains a C-terminal region similar to the terminal region 5' of the clones G+E+A of MSRV-1, which allows one to define the C-terminal region of the gene gag and a new region corresponding to the N-terminal region of the gene gag of MSRV-1.

CL2 allows one to define a region of 1511 pb that has an open phase of reading in the N-terminal region of 1077 pb that codes for 359 amino acids and one non-open phase of reading, of 454 pb, which corresponds to the C-terminal region of the gene gag of MSRV-1.

The nucleotide sequence of CL2 is identified by SEQ ID No: 130. It is represented in figure XX3.1, with the potential amino acid reading frames.

The fragment of 1077 pb of CL2 that codes for 359 amino acids was amplified by PCR with the enzyme Pwo (5U/microliter) (Boehringer Mannheim, France) using 1 microliter of the DNA minipreparation of the clone 2 under the following conditions: 95 °C for 1 minute, 60 °C for 1 minute, 72 °C for 2 minutes during 25 cycles and with a final reaction volume of 50 μl by the help of /28 the primers:

- Primer 5' (Bam HI), identified by SEQ ID No: 132
- 5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), and
- Primer 3' (Hind III), identified by SEQ ID No: 133
- 5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer) that correspond, respectively, to the nucleotide sequence of the clone 2 in position -9 to 21 and 1066 to 1095.

The fragment obtained after PCR was straightened out by Bam HI and HindIII and sub-cloned in the expression vectors pET28C and pET21C (Novagen) straightened by Bam HI and Hind III. The DNA sequencing of the fragment of 1077 pb of the clone 2 in the two expression vectors was carried out according to the method recommended for the use of the sequencing kit "PRISM® Ready Reaction Amplitaw® FS, DyeDeoxy® Terminator" (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the devices 373 A and 377 Applied Biosystems, according to the instruction of the manufacturer.

The expression of the nucleotide sequence of the fragment of 1077 pb of the clone 2 by the expression vectors pET28C and pET21C are identified by SEQ ID NO: 135 and SEQ ID NO: 137 respectively.

EXAMPLE 6: EXPRESSION OF CLONE 2 IN ESCHERICHIA COLI

The constructions pET28c-clone 2 (1077 pb) and pET21C-clone 2 (1077 pb) synthesize, in the bacterial strain BL21 (DE3), a protein with N- and C-terminal fusion for the vector pER28C and C-terminal for the vector pET21C with 6 histidines, with apparent molecular weight of about 45 kDa, demonstrated by polyacrylamide gel electrophoresis SDS-PAGE (SDA = Docecyl sodium sulfate) (Laemmli, 1970 (1). The reactivity of the protein was demonstrated vis-à-vis an anti-Histidine monoclonal antibody (Dianova) by the Western /29 blot technique (Towbin, et al., 1979 (2)).

The recombinant proteins pET28C-clone 2 (1077 pb) and pET21C-clone 2 (1077 pb) were displayed in SDS-PAGE in the insoluble fraction after enzymatic digestion of the bacterial extracts with 50 μl of lysozyme (10 mg/ml) and ultrasound lysis.

The antigen properties of the recombinant antigens pET28C-clone 2 (1077 pb) and pET21C-clone 2 (1077 pb) were tested by Western Blot technique () after solubilization of the bacterial residue with 2% SDS and 50 mM of beta-mercaptoethanol. After incubation with the sera of patients afflicted with multiple sclerosis, the sera of the neurological controls and the reference sera of the blood transfusion center (CTS), the immunocomplexes were detected by the help of anti-IgG goat serum and human anti-AgM, coupled with alkaline phosphatase.

The results are presented in the following table.

TABLE

Reactivity of sera afflicted with multiple sclerosis and references with the recombinant protein MSRV-1 gag clone 2 (1077 pb) = pET21C-clone 2 (1077 pb) and pET28C-clone 2 (1077 pb)[a]

Disease	Number of Tested individuals	Number of Positive Individuals
MS	15	6 2(+++),2(++), 2(+)
Neurological References	2	1(+++)
Healthy References (CTS)	22	1(+/-)

a) The small bands that contain 1.5 microgram of recombinant antigen pET-gag clone 2 have a reactivity against sera diluted to 1/100. The Western Blot interpretation is based on the presence or the absence of a band pET-gag clone 2 (1077 pb) specific on the bands. Some positive and negative controls are included in each experiment. /30

These results show that, under the technical conditions used, about 40% of the human sera afflicted with multiple sclerosis that were tested react with the recombinant proteins pET28C-clone 2 (1077 pb) and pET21C-clone 2 (1077 pb). A reactivity was observed for a neurological reference and it is interesting to note that the RNA extracted from this serum, after the reverse transcriptase stage, are also amplified by PCR in the pol region. This suggests that persons who have not been declared as having MS can also harbor and express this virus. On the other hand, an apparently healthy reference sample (CTS donor) has some anti-gag (clone 2, 1077 pb) antibodies. This is compatible with acquired immunity against MSRV-1 in addition to a declared associated auto-immune disease.

EXAMPLE 7: PRODUCTION OF A CLONE LB13 THAT CONTAINS AT 3' ONE PART HOMOLOGOUS TO THE CLONE 2 CORESPONDING TO THE GENE GAG AND AT 5' ONE PART HOMOLOGOUS TO THE CLONE 5M6 CORRESPONDING TO THE REGION LTR U5.

One RT-PCR ("reverse transcriptase polymerase chain reaction) was carried out from the total RNA extracted from virions that came from surface fluids of lymph B cells of patients afflicted with multiple sclerosis, concentrated by ultra centrifugings. The synthesis of cDNA was carried out with a specific primer SEQ No XXX and the reverse transcriptase "Expand® RT" of Boehringer Mannheiv according to the conditions recommended by the company.

Primer used for the synthesis of the cDNA, identified by SEQ /31 ID No: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

An amplification by PCR was carried out with the Taq polymerase (Perkin Elmer, France) under the following conditions: 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 2 minutes during 35 cycles and 72 °C for 7 minutes and with a final reaction volume of 100 μl.

Primers used for the amplification by PCR:

- Primer 5', identified by SEQ ID No: 139
- 5' TGT CCG CTG TGC TCC TGA TC 3'
- Primer 3', identified by SEQ ID No: 138
- 5' CTT GGA GGG TGC ATA ACC AGG CAA T 3'

A second so-called "semi-strain" amplification by PCR was carried out with a primer 3' located inside the region already amplified. This second amplification was carried out under the same experimental conditions as those used during the first amplification, using 10 μl of the amplification product derived from the first PCR.

Primers used for the amplification by "semi-strain" PCR:

- Primer 5', identified by SEQ ID No: 139
- 5' TGT CCG CTG TGC TCC TGA TC 3'
- Primer 3', identified by SEQ ID No: 140
- 5' CTA TGT CCT TTT GGA CTG TTT GGG T3'

The primers SEQ ID No: 138 and SEQ ID NO: 140 are specific for the region gag, clone 2 nucleotide position No. 373-397 and No. 433-456. The primers used in the region 5' were defined on some sequences of clones obtained during prior tests.

An amplification product of 764 pb was obtained and cloned in the following way:

The amplified DNA was inserted in a plasmid by the help of the TA Cloning® kit. Two μl of DNA solution were mixed with 5 μl of

sterile distilled water, 1 μl of a ligation buffer concentrated ten times "10X Ligation Buffer," 2 μl of "pCR® VECTOR" (25 ng/ml) and /32
1 μl of "T4 DNA Ligase." This mixture was incubated overnight at 14 °C. The following stages were carried out in conformity with the instructions of the TA Cloning® kit (Invitrogen). The mixture was spread out after transformation of the ligation is some E. coli bacteria INVαF'. At the end of the procedure, the white colonies of recombinant bacteria were subcultured to be cultivated and to allow the extraction of the plasmids incorporated according to the procedure called "mini-preparation of DNA" (17). The plasmid preparation of each recombinant colony was cut by the restriction enzyme Eco RI and analyzed on agar gel. The plasmids that have an insert detected under ultra-violet light after marking of the gel with ethidium bromide were selected for the sequencing of the insert, following hybridization with a complementary primer of the promoter T7 present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for the use of the sequencing kit "PRISM® Ready Reaction Amplitaq® FS, DyeDeoxy® Terminator" (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the devices 373 A and 377 Applied Biosystems, according to the instructions of the manufacturer.

The resulting clone LB13 contains a N-terminal region of the gene gag MSRV-1 homologous to the clone 2 and an LTR corresponding to one part of the U5 region. Between the U5 region and gag a fixation site for the transfer RNA, the PBS "primer binding site" was identified.

The nucleotide sequence of the fragment of 764 pb of the clone LB13 in the plasmid "pCR® vector" is represented in the identifier SEQ ID No: 141.

The fixation site for the transfer RNA, which has a sequence of the PBS tryptophan type, was identified in the nucleotide /33
position No. 342-359 of the clone LB13.

Another clone, named LA15, was obtained on the total RNA extracted from virions concentrated by ultra-centrifuging from a culture surface fluid of synoviocytes derived from a patient afflicted with rheumatoid arthritis. The strategy of amplification and cloning of the clone LA15 is exactly the same that was used for the clone LB13.

The nucleotide sequence of the clone LA15 that is represented in the identifier SEQ ID No: 142 is very similar to the clone LB13. This suggests that the retrovirus MSVR-1 detected in multiple sclerosis has some sequences similar to those encountered in

rheumatoid arthritis.

REFERENCES /34

(1) Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage t4. Nature. (1970). 227: 680-685.

(2) Towbin H., Staehelin T. and Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA. (1979). 76: 4350-4354.

[LIST OF SEQUENCES]
Pp. 35-47 are not translated as indicated on the original.

CLAIMS /48

1. Nuclear material, in the isolated or purified state, which includes a nucleotide sequence chosen in the group that consists of (i) the sequences SEQ ID No: 112, SEQ ID No: 114, SEQ ID No: 117, SEQ ID No: 120, SEQ ID No: 124, SEQ ID No: 130, SEQ ID No: 141, and SEQ ID No: 142; (ii) the complementary sequences of the sequences (i); and (iii) the sequences equivalent to the sequences (i) or (ii), in particular the sequences that have for all of their 100 contiguous monomers, at least 50%, and preferably at least 70% homology with the sequences (i) or (ii) respectively.

2. Nuclear material, in the isolated or purified state, which codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with a peptide sequence chosen in the group that consists of SEQ ID No: 113, SEQ ID No: 115, SEQ ID No: 118, SEQ ID No: 121, SEQ ID No: 135, and SEQ ID No: 137.

3. Retroviral nuclear material in which the gene pol includes a nucleotide sequence identical to or equivalent to a sequence chosen in the group that consists of SEQ ID No: 112, SEQ ID No: 124, and their complementary sequences.

4. Retroviral nuclear material in which the end 5′ of the gene pol begins at the nucleotide 1419 of the SEQ ID No: 130.

5. Retroviral nuclear material in which the gene pol codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 505, and preferably at least 70% homology with the peptide sequence SEQ ID No: 113.

6. Retroviral nuclear material in which the end 3′ of the gene gag ends at the nucleotide 1418 of the SEQ ID No; 130.

7. Retroviral nuclear material in which the gene env includes /49 a nucleotide sequence identical to or equivalent to a sequence chosen in the group that consists of SEQ ID No: 117, and its complementary sequences.

8. Retroviral nuclear material in which the gene env includes a nucleotide sequence that begins at the nucleotide 1 of SEQ ID NO: 117 and ends at the nucleotide 233 of SEQ ID No: 114.

9. Retroviral nuclear material in which the gene env codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50% and preferably at least 70%

homology with the sequence SEQ ID No: 118.

10. Retroviral nuclear material in which the region U3R of LTR 3′ includes a nucleotide sequence that terminates at nucleotide 617 of SEQ ID No: 114.

11. Retroviral nuclear material in which the region RU5 of LTR 5′ includes a nucleotide sequence that begins at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID No: 141 or SEQ ID No: 142.

12. Retroviral nuclear material that includes a sequence that begins at nucleotide 755 of SEQ ID No: 120 and that terminates at nucleotide 617 of SEQ ID No: 114.

13. Retroviral nuclear material according to any of the preceding claims characterized in that it is associated with at least one auto-immune disease such as multiple sclerosis or rheumatoid arthritis.

14. Nucleotide fragment that includes a nucleotide sequence chosen in the group that consists of (i) the sequences SEQ ID No: 112, SEQ ID No: 114, SEQ ID No: 117, SEQ ID No: 120, SEQ ID No: 124, SEQ ID No: 130, SEQ ID No: 141, and SEQ ID No: 142; (ii) the complementary sequences of the sequences (i); and (iii) the sequences equivalent to the sequences (i) or (ii), in particular the sequences that have for the entire series of 100 contiguous monomers, at least 50%, and preferably at least 70% homology with the sequences (i) or (ii) respectively. /50

15. Nucleotide fragment according to Claim 14 consisting of a nucleotide sequence chosen in the group that consists of (i) the sequences SEQ ID No: 112, SEQ ID No: 114, SEQ ID No: 117, SEQ ID No: 120, SEQ ID No: 124, SEQ ID No: 130, SEQ ID No: 141, and SEQ ID No: 142; (ii) the complementary sequences of the sequences (i); and (iii) the sequences equivalent to the sequences (i) or (ii), in particular the sequences that have for the entire series of 100 contiguous monomers, at least 50%, and preferably at least 70% homology with the sequences (i) or (ii) respectively.

16. Nucleotide fragment that includes a nucleotide sequence that codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with a peptide sequence chosen in the group that consists of SEQ ID No: 113, SEQ ID No: 115, SEQ ID No: 118, SEQ ID No: 121, SEQ ID No: 135, and SEQ ID No: 137.

17. Nucleotide fragment according to Claim 16 consisting in a nucleotide sequence that codes for a polypeptide that has, for the entire contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology, with a peptide sequence chosen in the group that consists of SEQ ID No: 113, SEQ ID No: 115, SEQ ID No: 118, SEQ ID No: 121, SEQ ID No: 135, and SEQ ID No: 137.

18. Nuclear probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, characterized in that it is capable of being hybridized specifically on any fragment according to any of Claims 14 to 17, which belong to /51 the genome of the said retrovirus.

19. Probe according to Claim 18 characterized in that it has from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides.

20. Primer for the amplification by polymerization of RNA or DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis characterized in that it includes a nucleotide sequence identical or equivalent to at least one part of the nucleotide sequence of a fragment according to any one of Claims 8 to 11, especially a nucleotide sequence that has for any series of 10 contiguous monomers, at least 50%, and preferably at least 70% homology with at least the said part of the said fragment.

21. Primer according to Claim 20 characterized in that its nucleotide sequence is chosen among SEQ ID No: 116, SEQ ID No: 119, SEQ ID No: 122, SEQ ID No: 123, SEQ ID No: 126, SEQ ID No: 127, SEQ ID No: 128, SEQ ID No: 129, SEQ ID No: 132, and SEQ ID No: 133.

22. RNA or DNA, and especially the replication and/or expression vector, which includes a genomic fragment of the nuclear material according to any of Claims 1 to 7 or a fragment according to any of Claims 14 to 17.

23. Peptide coded by any open reading frame that belongs to a nucleotide fragment according to any of Claims 14 to 17, especially a polypeptide, an oligopeptide for example that forms or includes an antigen determinant recognized by the sera of patients infected by the virus MSRV-1, and/or in which the virus MSRV-1 and been reactivated.

24. Peptide according to Claim 23 that includes a sequence identical to, partially or completely, or equivalent to a sequence chosen among SEQ ID No: 113, SEQ ID No: 115, SEQ ID No: 118, SEQ /52 ID No: 121, SEQ ID No: 135, and SEQ ID No: 137.

25. Prophylactic or therapeutic diagnostic compound, especially for inhibiting the expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, which includes a nucleotide fragment according to any of Claims 14 to 17.

26. Process for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, characterized in that one puts in contact an RNA and/or a DNA presumed to belong to or come from the said retrovirus, or their complementary RNA and/or DNA, with a compound that includes a nucleotide fragment according to any of Claims 14 to 17.

[Figure 1]
Key:
AND PROVIRAL=retroviral DNA;
ARN GENOMIQUE (VIRION)=genomic RNA (virion).

[Figures 2] to [Figure 11]
Lists of gene sequences;
(suite)=(continuation).

FIG 1

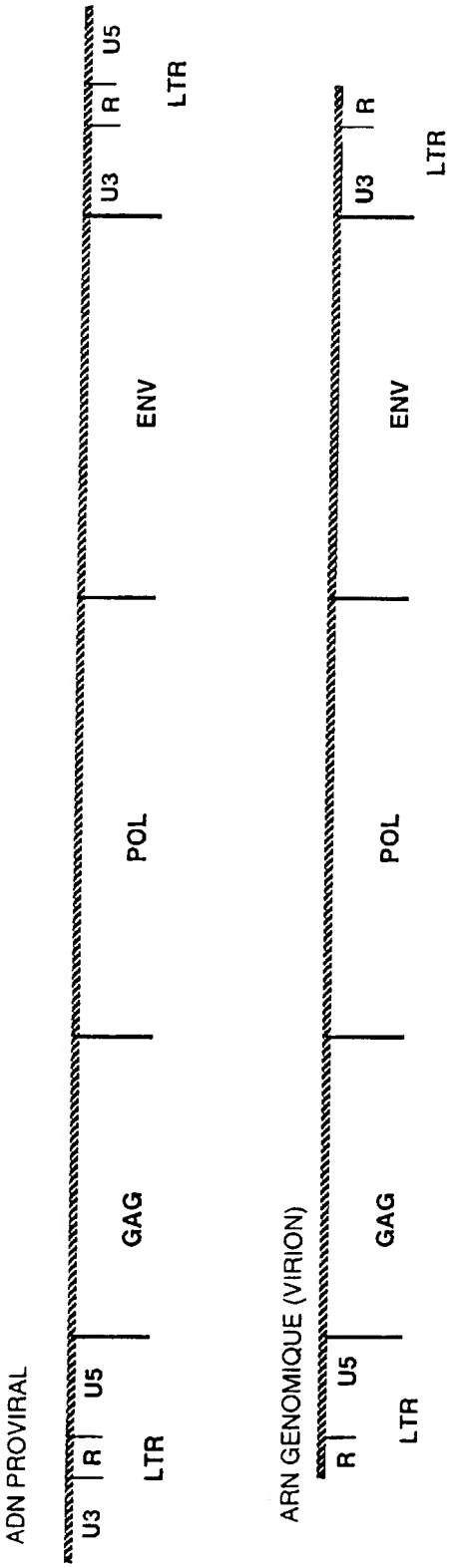
ADN PROVIRAL
U3
R
U5
LTR
GAG
POL
ENV
U3
R
U5
LTR
ARN GENOMIQUE (VIRION)
R
U5
LTR
GAG
POL
ENV
U3
R
LTR

FIG 2

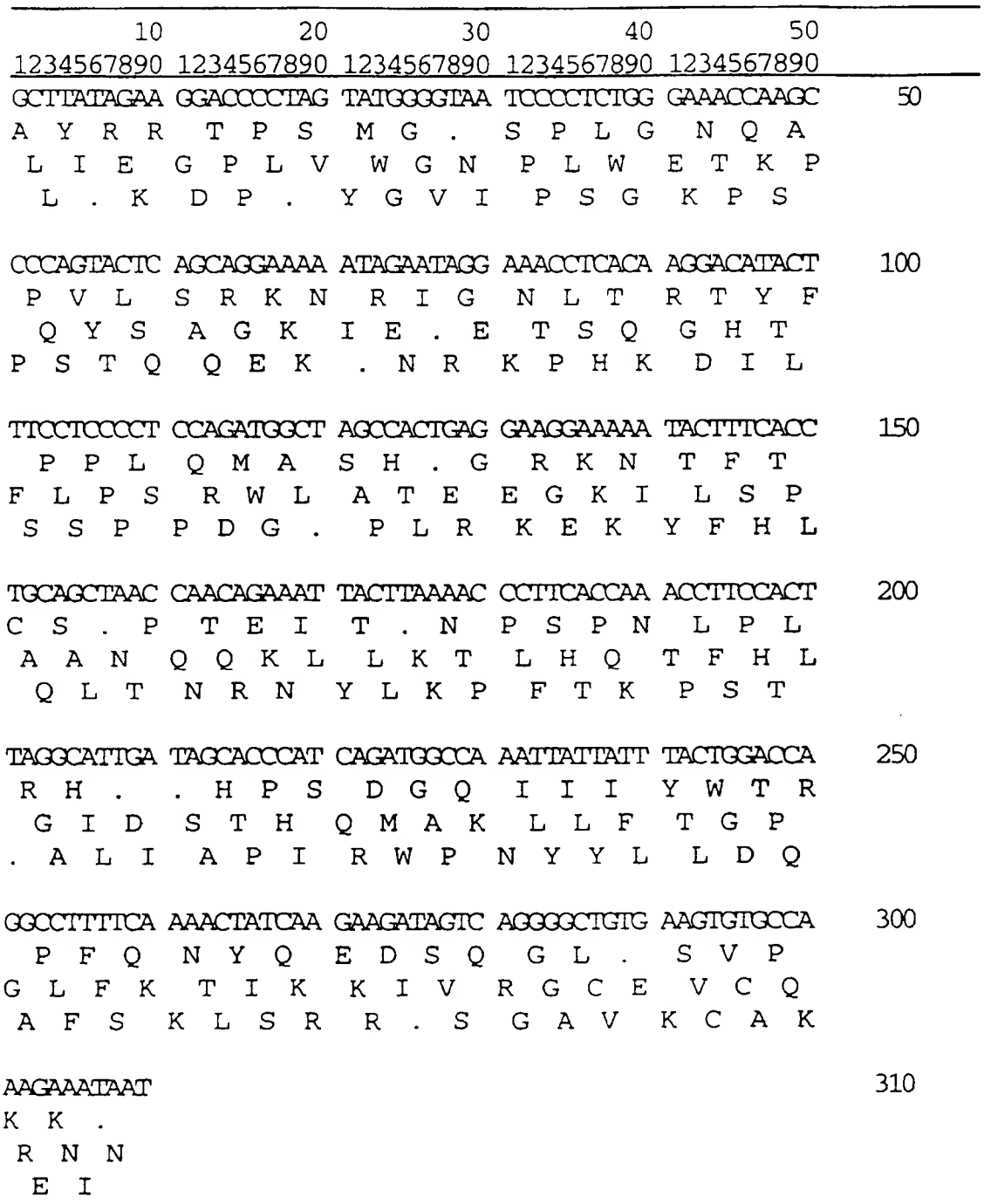
10 20 30 40 50
1234567890 1234567890 1234567890 1234567890 1234567890
GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC 50
A Y R R T P S M G . S P L G N Q A
L I E G P L V W G N P L W E T K P
L . K D P . Y G V I P S G K P S
CCCAGTACTC AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT 100
P V L S R K N R I G N L T R T Y F
Q Y S A G K I E . E T S Q G H T
P S T Q Q E K . N R K P H K D I L
TTCCTCCCCT CCAGATGGCT AGCCACTGAG GAAGGAAAAA TACTTTCACC 150
P P L Q M A S H . G R K N T F T
F L P S R W L A T E E G K I L S P
S S P P D G . P L R K E K Y F H L
TGCAGCTAAC CAACAGAAAT TACTTAAAAC CCTTCACCAA ACCTTCCACT 200
C S . P T E I T . N P S P N L P L
A A N Q Q K L L K T L H Q T F H L
Q L T N R N Y L K P F T K P S T
TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT TACTGGACCA 250
R H . . H P S D G Q I I I Y W T R
G I D S T H Q M A K L L F T G P
. A L I A P I R W P N Y Y L L D Q
GGCCTTTTCA AAACTATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300
P F Q N Y Q E D S Q G L . S V P
G L F K T I K K I V R G C E V C Q
A F S K L S R R . S G A V K C A K
AAGAAATAAT 310
K K .
R N N
E I

FIG 2 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC     50
P  C  I  F     N  L  L     V  K  F     V  S  S  R     I  K  T
 P  V  S     L  T  S  L     L  S  L     S  L  P     E  S  K  L
  L  Y  L     .  P  P     C  .  V  C     L  F  Q     N  Q  N

TGTAAAACTA CAAATTGTTC TTCAAATGGA GCACCAGATG GAGTCCATGA    100
 V  K  L     Q  I  V  L     Q  M  E     H  Q  M     E  S  M  T
  .  N  Y     K  L  F     F  K  W  S     T  R  W     S  P  .
C  K  T  T     N  C  S     S  N  G     A  P  D  G     V  H  D

CTAAGATCCA CCGTGGACCC CTGGACCGGC CTGCTAGCCC ATGCTCCGAT    150
  K  I  H     R  G  P     L  D  R  P     A  S  P     C  S  D
L  R  S  T     V  D  P     W  T  G     L  L  A  H     A  P  M
 .  D  P     P  W  T  P     G  P  A     C  .  P     M  L  R  C

GTTAATGACA TTGAAGGCAC CCCTCCCGAG GAAATCTCAA CTGCACAACC    200
V  N  D  I     E  G  T     P  P  E     E  I  S  T     A  Q  P
 L  M  T     L  K  A  P     L  P  R     K  S  Q     L  H  N  P
  .  .  H     .  R  H     P  S  R  G     N  L  N     C  T  T

CCTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCGGT CATCAGCCAA    250
 L  L  C     P  N  S  A     G  S  S     .  S  G     H  Q  P  T
  Y  Y  A     P  I  Q     R  E  A  V     R  A  V     I  S  Q
P  T  M  P     Q  F  S     G  K  Q     L  E  R  S     S  A  N

CCTCCCCAAC AGCACTTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC    300
  S  P  T     A  L  G     F  S  C  .     E  G  G     L  R  D
P  P  Q  Q     H  L  G     F  P  V     E  R  G  D     .  E  T
 L  P  N     S  T  W  V     F  L  L     R  G  G     T  E  R  Q

AGGACTAGCT GGATTTCCTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG    350
R  T  S  W     I  S  .     A  N  E     E  S  L  S     L  A  G
 G  L  A     G  F  P  R     P  T  K     N  P  .     A  .  L  G
  D  .  L     D  F  L     G  Q  R  R     I  P  K     P  S  W
```

FIG 3

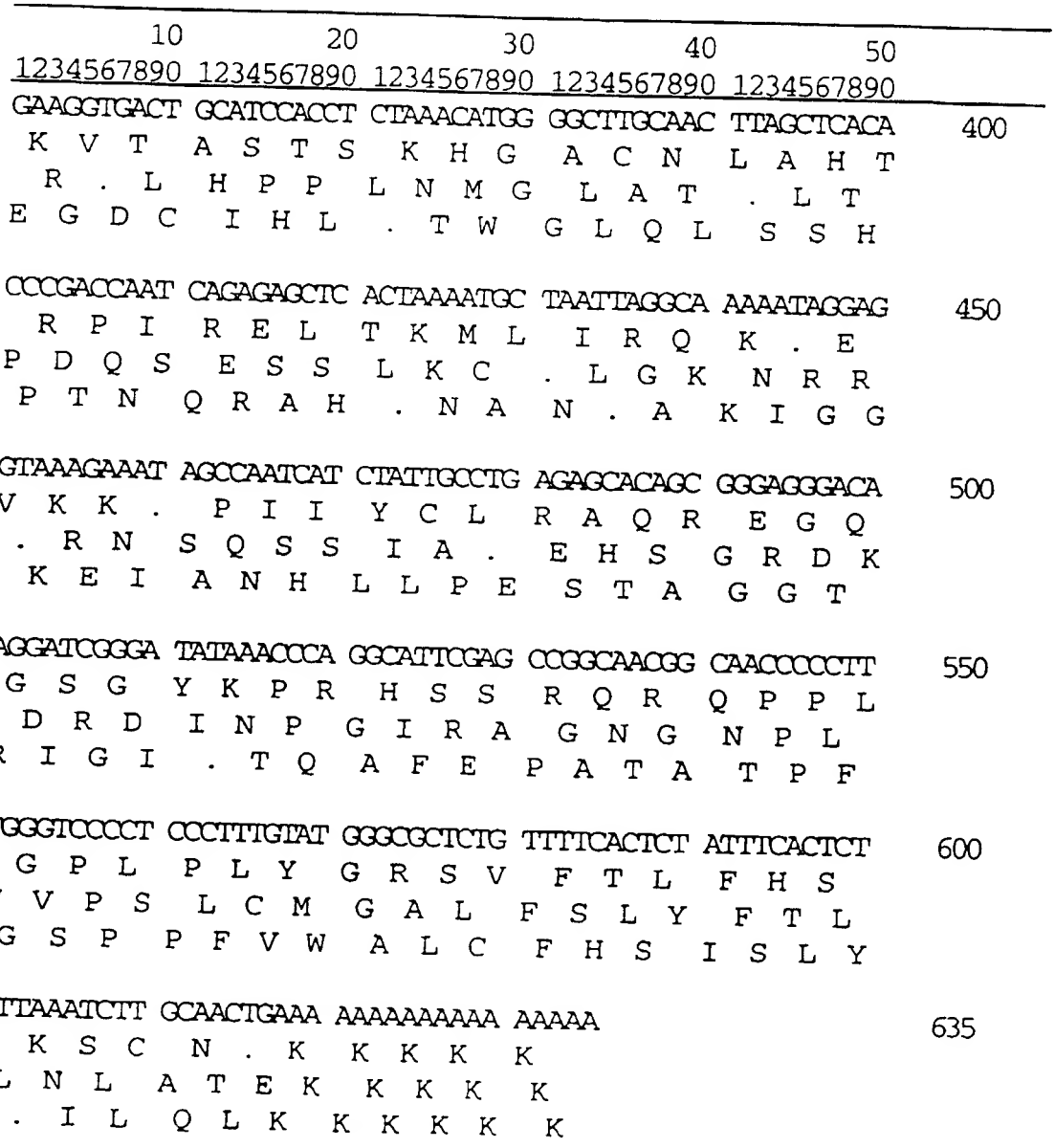
10 20 30 40 50
1234567890 1234567890 1234567890 1234567890 1234567890
GAAGGTGACT GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA 400
K V T A S T S K H G A C N L A H T
R . L H P P L N M G L A T . L T
E G D C I H L . T W G L Q L S S H
CCCGACCAAT CAGAGAGCTC ACTAAAATGC TAATTAGGCA AAAATAGGAG 450
R P I R E L T K M L I R Q K . E
P D Q S E S S L K C . L G K N R R
P T N Q R A H . N A N . A K I G G
GTAAAGAAAT AGCCAATCAT CTATTGCCTG AGAGCACAGC GGGAGGGACA 500
V K K . P I I Y C L R A Q R E G Q
. R N S Q S S I A . E H S G R D K
K E I A N H L L P E S T A G G T
AGGATCGGGA TATAAACCCA GGCATTCGAG CCGGCAACGG CAACCCCCTT 550
G S G Y K P R H S S R Q R Q P P L
D R D I N P G I R A G N G N P L
R I G I . T Q A F E P A T A T P F
TGGGTCCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTTCACTCT 600
G P L P L Y G R S V F T L F H S
W V P S L C M G A L F S L Y F T L
G S P P F V W A L C F H S I S L Y
ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635
I K S C N . K K K K K
L N L A T E K K K K K
. I L Q L K K K K K K

FIG 4

```
      10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCCTTT   50
M  A  L  P  Y  H  T  F  L  F  T  V  L  L  P  P  F
 W  P  S  L  I  I  L  F  S  L  L  F  S  Y  P  L  S
  G  P  P  L  S  Y  F  S  L  Y  C  S  L  T  P  F

CGCTCTCACT GCACCCCCTC CATGCTGCTG TACAACCAGT AGCTCCCCTT  100
 A  L  T  A  P  P  P  C  C  C  T  T  S  S  S  P  Y
  L  S  L  H  P  L  H  A  A  V  Q  P  V  A  P  L
R  S  H  C  T  P  S  M  L  L  Y  N  Q  .  L  P  L

ACCAAGAGTT TCTATGAAGA ACGCGGCTTC CTGGAAATAT TGATGCCCCA  150
  Q  E  F  L  .  R  T  R  L  P  G  N  I  D  A  P
T  K  S  F  Y  E  E  R  G  F  L  E  I  L  M  P  H
 P  R  V  S  M  K  N  A  A  S  W  K  Y  .  C  P  I

TCATATAGGA GTTTATCTAA GGGAAACTCC ACCTTCACTG CCCACACCCA  200
S  Y  R  S  L  S  K  G  N  S  T  F  T  A  H  T  H
 H  I  G  V  Y  L  R  E  T  P  P  S  L  P  T  P  I
  I  .  E  F  I  .  G  K  L  H  L  H  C  P  H  P

TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG CATGCAAATA  250
 M  P  R  N  C  Y  N  S  A  T  L  C  M  H  A  N  T
  C  P  A  T  A  I  T  L  P  L  F  A  C  M  Q  I
Y  A  P  Q  L  L  .  L  C  H  S  L  H  A  C  K  Y

CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT  300
  H  Y  W  T  G  K  M  I  N  P  S  C  P  G  G  L
L  I  I  G  Q  G  K  .  L  I  L  V  V  L  E  D  L
 S  L  L  D  R  E  N  D  .  S  .  L  S  W  R  T  W

GGAGCCACTG TCTGTTGGAC TTACTTCACC CATACCAGTA TGTCTGATGG  350
G  A  T  V  C  W  T  Y  F  T  H  T  S  M  S  D  G
 E  P  L  S  V  G  L  T  S  P  I  P  V  C  L  M  G
  S  H  C  L  L  D  L  L  H  P  Y  Q  Y  V  .  W
```

FIG 4 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ACCTCACCTG TGTAAAATTT AGCAATACTA TAGACACAAC CAGCTCCCAA   750
  L T C  V K F  S N T I  D T T  S S Q
T S P V  . N L  A I L  . T Q P  A P N
 P H L  C K I  .  Q Y Y  R H N  Q L P M

TGCATCAGGT GGGTAACACC TCCCACACGA ATAGTCTGCC TACCCTCAGG   800
C I R W  V T P  P T R  I V C L  P S G
 A S G  G . H L  P H E  . S A  Y P Q E
  H Q V  G N T  S H T N  S L P  T L R

AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG AATGGCTCTT   850
 I F F  V C G T  S A Y  H C L  N G S S
  Y F L  S V V  P Q P I  I V .  M A L
N I F C  L W Y  L S L  S L F E  W L F

CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC   900
  E S M  C F L  S F L V  P P M  T I Y
Q N L C  A S S  H S .  C P L .  P S T
 R I Y  V L P  L I L S  A P Y  D H L H

ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG   950
T E Q D  L Y N  H V V  P K P H  N K R
 L N K  I Y T I  M S Y  L S P  T T K E
  . T R  F I Q  S C R T  . A P  Q Q K

AGTACCCATT CTTCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG  1000
 V P I  L P F V  I R A  G V L  G R L G
  Y P F  F L L  L S E Q  E C .  A D .
S T H S  S F C  Y Q S  R S A R  Q T R

GTACTGGCAT TGGCAGTATC ACAACCTCTA CTCAGTTCTA CTACAAACTA  1050
  T G I  G S I  T T S T  Q F Y  Y K L
V L A L  A V S  Q P L  L S S T  T N Y
 Y W H  W Q Y H  N L Y  S V L  L Q T I
```

FIG 4 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
TCTCAAGAAA TAAATGGTGA CATGGAACAG GTCACTGACT CCCTGGTCAC   1100
S Q E I  N G D  M E Q  V T D S  L V T
 L K K  . M V T  W N R  S L T  P W S P
  S R N  K W .  H G T G  H . L  P G H

CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT CAAAATCGAA   1150
 L Q D  Q L N S  L A A  V V L  Q N R R
  C K I  N L T  P . Q Q  . S F  K I E
L A R S  T . L  P S S  S S P S  K S K

GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTTA   1200
  A L D  L L T  A K R G  G T C  L F L
E L . T  C . P  P K E  G E P V  Y F .
 S F R  L A N R  Q K R  G N L  F I F R

GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA   1250
G E E R  C Y Y  V N Q  S R I V  T E K
 E K N  A V I M  L I N  P E L  S L R K
  R R T  L L L  C . S I  Q N C  H . E

AGTTAAAGAA ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAAA   1300
 V K E  I R D R  I Q C  R A E  E L Q N
  L K K  F E I  E Y N V  E Q R  S F K
S . R N  S R S  N T M  . S R G  A S K

ACACCGAACG CTGGGGCCTC CTCAGCCAAT GGATGCCCTG GGTTCTCCCC   1350
  T E R  W G L  L S Q W  M P W  V L P
T P N A  G A S  S A N  G C P G  F S P
 H R T  L G P P  Q P M  D A L  G S P L

TTCTTAGGAC CTCTAGCAGC TCTAATATTG TTACTCCTCT TTGGACCCTG   1400
F L G P  L A A  L I L  L L L F  G P C
 S . D  L . Q L  . Y C  Y S S  L D P V
  L R T  S S S  S N I V  T P L  W T L
```

FIG 4 (suite)

```
         10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT GAAGCTGTAA   1450
 I  F  N   L  L  V  K   F  V  S   S  R  I   E  A  V  K
  S  L  T   S  L  L   S  L  S  L   P  E  L   K  L  .
Y  L  .  P   P  C  .   V  C  L   F  Q  N  .   S  C  K

AGCTACAGAT GGTCTTACAA ATGGAACCCC A                      1481
  L  Q  M   V  L  Q   M  E  P
S  Y  R  W   S  Y  K   W  N  P
 A  T  D   G  L  T  N   G  T  P
```

FIG 5

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT      50
S K S K  S F R  L A N  R Q K R  G N L
 Q N R  R A L D  L L T  A K R  G G T C
  K I E  E L .  T C .  P  P K E  G E P

GTTTATTTTT AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC     100
 F I F  R G R M  L L V  C .  S  I W N H
  L F L  G E E  C C .  Y  V N Q  S G I
V Y F .  G K N  A V S  M L I N  L E S

ATTACTGAGA AAGTTAAAGA AATTTGAGAT CGAATATAAT GTAGAGCAGA     150
  Y .  E  S .  R  N L R S  N I M  .  S R
I T E K  V K E  I .  D  R I .  C  R A E
 L L R  K L K K  F E I  E Y N  V E Q R

GGACCTTCAA AACACTGCAC CCTGGGGCCT CCTCAGCCAA TGGATGCCCT     200
G P S K  H C T  L G P  P Q P M  D A L
 D L Q  N T A P  W G L  L S Q  W M P W
  T F K  T L H  P G A S  S A N  G C P

GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT TTTACTCCTC     250
 D S P  L L R T  S S S  Y N I  F T P L
  T L P  F L G  P L A A  I I F  L L L
G L S P  S .  D  L .  Q  L .  Y F  Y S S

TTTGGACCCT GTATCTTCAA CTTCCTTGTT AAGTTTGTCT CTTCCAGAAT     300
  W T L  Y L Q  L P C .  V C L  F Q N
F G P C  I F N  F L V  K F V S  S R I
 L D P  V S S T  S L L  S L S  L P E L

TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT     350
 .  S C K  A T N  S S S  N G T P  D A V
 E A V  K L Q I  V L Q  M E P  Q M Q S
  K L .  S Y K  .  F F K  W N P  R C S
```

FIG 5 (suite)

```
      10        20        30        40        50
1234567890 1234567890 1234567890 1234567890 1234567890
CCATGACTAA AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC   400
 H  D  .  N  L  P  W  T  P  G  P  A  C  .  T  M  L
  M  T  K  I  Y  R  G  P  L  D  R  P  A  R  L  C
P  .  L  K  S  T  V  D  P  W  T  G  L  L  D  Y  A

TCTGATGTTA ATGACATTGA AGTCACCCCT CCCGAGGAAA TCTCAACTGC   450
  .  C  .  .  H  .  S  H  P  S  R  G  N  L  N  C
S  D  V  N  D  I  E  V  T  P  P  E  E  I  S  T  A
 L  M  L  M  T  L  K  S  P  L  P  R  K  S  Q  L  H

ACAACCCCTA CTACACTCCA ATTCAGTAGG AAGCAGTTAG AGCAGTTGTC   500
T  T  P  T  T  L  Q  F  S  R  K  Q  L  E  Q  L  S
 Q  P  L  L  H  S  N  S  V  G  S  S  .  S  S  C  Q
  N  P  Y  Y  T  P  I  Q  .  E  A  V  R  A  V  V

AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA GGGTGGACTG   550
 A  N  L  P  N  S  T  W  V  F  L  L  R  G  W  T  E
  P  T  S  P  T  V  L  G  F  S  C  .  E  G  G  L
S  Q  P  P  Q  Q  Y  L  G  F  P  V  E  R  V  D  .

AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT   600
  R  Q  D  .  L  D  F  L  G  .  L  R  I  P  K  P
R  D  R  T  S  W  I  S  .  A  D  .  E  S  X  S  L
 E  T  G  L  A  G  F  P  R  L  T  K  N  P  X  A  X

ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG   650
X  W  E  G  D  R  I  H  L  .  T  W  G  L  Q  L  S
 X  G  K  V  T  A  S  I  F  K  H  G  A  C  N  L  A
  L  G  R  .  P  H  P  S  L  N  M  G  L  A  T  .

CTCACACCCG ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA   700
 S  H  P  T  N  Q  R  A  H  .  N  A  N  Q  A  K  T
  H  T  R  P  I  R  E  L  T  K  M  L  I  R  Q  K
L  T  P  D  Q  S  E  S  S  L  K  C  .  S  G  K  N
```

FIG 5 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CAGGAGGTAA AGCAATAGCC AATCATCTAT TGCCTGAGAG CACAGCGGGA   750
  G G K   A I A   N H L L   P E S   T A G
Q E V K   Q . P   I I Y   C L R A   Q R E
 R R .   S N S Q   S S I   A . E   H S G K

AGGACAAGGA TTGGGATATA AACTCAGGCA TTCAAGCCAG CAACAGCAAC   800
R T R I   G I .   T Q A   F K P A   T A T
 G Q G   L G Y K   L R H   S S Q   Q Q Q P
  D K D   W D I   N S G I   Q A S   N S N

CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT CACTCTATTT   850
 P F G   S P P I   V W E   L C F   H S I S
  P L G   P L P   L Y G S   S V F   T L F
P L W V   P S H   C M G   A L F S   L Y F

CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG   900
  L Y .   I M Q   L H S S   G P C   F L W
H S I K   S C N   C T L   L V R V   F Y G
 T L L   N H A T   A L F   W S V   F F M A

CTCAAGCTGA GCTTTTGTTC GCCATCCACC ACTGCTGTTT GCCACCGTCA   950
L K L S   F C S   P S T   T A V C   H R H
 S S .   A F V R   H P P   L L F   A T V T
  Q A E   L L F   A I H H   C C L   P P S

CAGACCCGCT GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG   1000
 R P A   A D F H   P F G   S S R   V S T V
  D P L   L T S   I P L D   P A E   C P L
Q T R C   . L P   S L W   I Q Q S   V H C

TGCTCCTGAT CCAGCGAGGT ACCCATTGCC ACTCCCGATC AGGCTAAAGG   1050
  L L I   Q R G   T H C H   S R S   G . R
C S .   S S E V   P I A   T P D Q   A K G
 A P D   P A R Y   P L P   L P I   R L K A
```

FIG 5 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CTTGCCATTG TTCCTGCATG GCTAAGTGCC TGGGTTTGTC CTAATAGAAC  1100
L A I V  P A W  L S A  W V C P  N R T
 L P L  F L H G  . V P  G F V  L I E L
  C H C  S C M  A K C L  G L S  . . N

TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC CCACGGCTTC  1150
 E H W  S L G S  M V L  F H D  P R L L
  N T G  H W V  P W F S  S M T  H G F
. T L V  T G F  H G S  L P .  P T A S

TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGGTA  1200
  I E L  . H S  P H G P  R F H  S L V
. . S Y  N T H  R M A  Q D S I  P W Y
 N R A  I T L T  A W P  K I P  F L G I

TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT  1250
S V R P  R T P  G Q R  X . G L  P P F
 L . G  Q E P Q  V R E  X E A  C H H L
  C E A  K N P  R S E X  V R L  A T I

TGGGAAGTGG CCCACTGCCA TTTTGGTAGC GGCCCACCAC CATCTTGGGA  1300
 G K W  P T A I  L V A  A H H  H L G S
  G S G  P L P  F W .  R P T T  I L G
W E V A  H C H  F G S  G P P P  S W E

GCTGTGGGAG CAAGGATCCC CCAGTAACA                         1329
  C G S  K D P  P V T
A V G A  R I P  Q .
 L W E  Q G S P  S N
```

FIG 6

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA     50
P  R  T  Y  S  G  E  L  G  P  M  .  H  S  D  A  K
 L  E  R  I  L  E  N  W  D  Q  C  D  T  Q  T  L  R
  .  N  V  F  W  R  I  G  T  N  V  T  L  R  R  .

GAAAGAAACG ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT    100
 K  E  T  I  Y  I  L  L  Q  Y  R  L  A  T  I  S  S
  K  K  R  F  I  F  F  C  S  T  A  W  P  Q  Y  P
E  R  N  D  L  Y  S  S  A  V  P  P  G  H  N  I  L

CTTCAAGGGA GAGAAACCTG GCTTCCTGAG GGAAGTATAA ATTATAACAT    150
  S  R  E  R  N  L  A  S  .  G  K  Y  K  L  .  H
L  Q  G  R  E  T  W  L  P  E  G  S  I  N  Y  N  I
 F  K  G  E  K  P  G  F  L  R  E  V  .  I  I  T  S

CATCTTACAG CTAGACCTCT TCTGTAGAAA GGAGGGCAAA TGGAGTGAAG    200
H  L  T  A  R  P  L  L  .  K  G  G  Q  M  E  .  S
 I  L  Q  L  D  L  F  C  R  K  E  G  K  W  S  E  V
  S  Y  S  .  T  S  S  V  E  R  R  A  N  G  V  K

TGCCATATGT GCAAACTTTC TTTTCATTAA GAGACAACTC ACAATTATGT    250
A  I  C  A  N  F  L  F  I  K  R  Q  L  T  I  M  .
 P  Y  V  Q  T  F  F  S  L  R  D  N  S  Q  L  C
C  H  M  C  K  L  S  F  H  .  E  T  T  H  N  Y  V

AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA    300
 K  V  W  F  M  P  Y  R  K  P  S  E  S  T  S  L
K  K  C  G  L  C  P  T  G  S  P  Q  S  P  P  P  Y
 K  S  V  V  Y  A  L  Q  E  A  L  R  V  H  L  P  T

CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT    350
P  Q  R  P  L  P  D  S  F  L  N  .  .  G  P  P  F
 P  S  V  P  S  P  T  P  S  S  T  N  K  D  P  P  L
  P  A  S  P  P  R  L  L  P  Q  L  I  R  T  P  L

TAACCCAAAC GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA    400
N  P  N  G  P  K  G  D  R  Q  R  G  K  Q  .  T  K
 T  Q  T  V  Q  K  E  I  D  K  G  V  N  N  E  P
.  P  K  R  S  K  R  R  .  T  K  G  .  T  M  N  Q

AAGAGTGCCA ATATTCCCCG ATTATGCCCC CTCCAAGCAG TGAGAGGAGG    450
  E  C  Q  Y  S  P  I  M  P  P  P  S  S  E  R  R
K  S  A  N  I  P  R  L  C  P  L  Q  A  V  R  G  G
 R  V  P  I  F  P  D  Y  A  P  S  K  Q  .  E  E  E

AGAATTCGGC CCAGCCAGAG TGCCTGTACC TTTTTCTCTC TCAGACTTAA    500
R  I  R  P  S  Q  S  A  C  T  F  F  S  L  R  L  K
 E  F  G  P  A  R  V  P  V  P  F  S  L  S  D  L  K
  N  S  A  Q  P  E  C  L  Y  L  F  L  S  Q  T  .
```

FIG 6 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
AGCAAATTAA AATAGACCTA GGTAAATTCT CAGATAACCC TGACGGCTAT   550
 A N . N R P R . I L R . P . R L Y
  Q I K I D L G K F S D N P D G Y
S K L K . T . V N S Q I T L T A I

ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA   600
  . C F T R V R T I L . S D M E R
I D V L Q G L G Q S F D L T W R D
 L M F Y K G . D N P L I . H G E I

TATAATGTTA CTACTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG   650
Y N V T T K S D T N P K . E K C R
 I M L L L N Q T L T P N E R S A A
  . C Y Y . I R H . P Q M R E V P

CTGTAACTGC AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC   700
 C N C S P R V W R S L V S Q S G Q
  V T A A R E F G D L W Y L S Q A
L . L Q P E S L A I F G I S V R P

AACAATAGGA TGACAACAGA GGAAAGAACA ACTCCCACAG GCCAGCAGGC   750
  Q . D D N R G K N N S H R P A G
N N R M T T E E R T T P T G Q Q A
 T I G . Q Q R K E Q L P Q A S R Q

AGTTCCCAGT GTAGACCCTC ATTGGGACAC AGAATCAGAA CATGGAGATT   800
S S Q C R P S L G H R I R T W R L
 V P S V D P H W D T E S E H G D W
  F P V . T L I G T Q N Q N M E I

GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT GAGGAAAACT   850
 V P Q T F A N L R A R R T E E N .
  C H K H L L T C V L E G L R K T
G A T N I C . L A C . K D . G K L

AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA   900
  E E A Y E L L N D V H Y N T G K
R K K P M N Y S M M S T I T Q G K
 G R S L . I T Q . C P L . H R E R

GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA   950
G R K S Y C F S G Q T K G G I E E
 E E N L T A F L D R L R E A L R K
  K K I L L L F W T D . G R H . G

AGCATACCTC CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG  1000
 A Y L P V T . L Y . R P T N L K G
  H T S L S P D S I E G Q L I L K
S I P P C H L T L L K A N . S . R
```

FIG 6 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
GATAAGTTTA TCACTCAGTC AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG  1050
  . V Y  H S V  S C R H  . K K  L Q K
D K F I  T Q S  A A D  I R K N  F K S
 I S L  S L S Q  L Q T  L E K  T S K V

TCTGCCTTAG GCCCGGAGCA GAACTTAGAA ACCCTATTTA ACTTGGCATC  1100
S A L G  P E Q  N L E  T L F N  L A S
 L P .  A R S R  T . K  P Y L  T W H P
  C L R  P G A  E L R N  P I .  L G I

CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA CGGGACAAAC  1150
 S V F  Y N R D  Q E E  Q A K  R D K R
  Q F F  I I E  I R R S  R R N  G T N
L S F L  . .  R S G G  A G E T  G Q T

GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC  1200
  D K K  K R G  G P L L  . S W  P S G
G I K K  K G G  V H Y  F S H G  P Q A
 G . K  K K G G  S T T  L V M  A L R Q

AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG  1250
K Q T L  E A L  Q K G  K A G Q  I K C
 S R L  W R L C  K R E  K L G  K S N A
  A D F  G G S  A K G K  S W A  N Q M

CCTAATAGGG CTGGCTTCCA GTGGGGTCTA CAAGGACACT TTAAAAAAGA  1300
 L I G  L A S S  A V Y  K D T  L K K I
  . . G  W L P  V R S T  R T L  . K R
P N R A  G F Q  C G L  Q G H F  K K D

TTATCCAAGT AGAAATAAGC CGCCCCCTTG TCCATGCCCC TTACGTCAAG  1350
  I Q V  E I S  R P L V  H A P  Y V K
L S K .  K . A  A P L  S M P L  T S R
 Y P S  R N K P  P P C  P C P  L R Q G

GGAATCACTG GAAGCCCCAC TGCCCCAGGG GATGAAGATA CTCTGAGTCA  1400
G I T G  R P T  A P G  D E D T  L S Q
 E S L  E G P L  P Q G  M K I  L . V R
  N H W  K A H  C P R G  . R Y  S E S

GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT GCCCGGGGCG  1450
 K P L  T R .  S S S R  T E G  A R G E
  S H .  P D D  P A A G  L R V  P G A
E A I N  Q M I  Q Q Q  D . G C  P G R

AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA  1500
  R Q P  M P S  P S Q S  P G Y  V . P
S A S P  C H H  P H R  A P G M  F D H
 A P A  H A I T  L T E  P R V  C L T I
```

FIG 6 (suite)

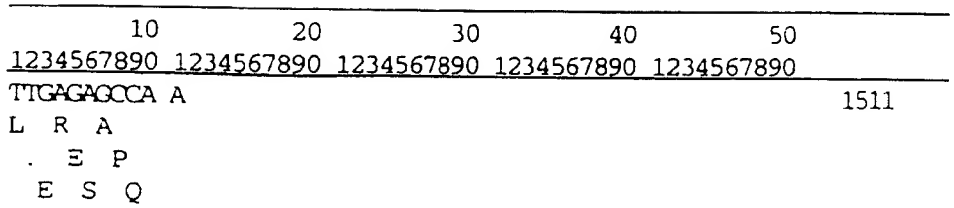
10 20 30 40 50
1234567890 1234567890 1234567890 1234567890 1234567890
TTGAGAGCCA A
1511
L R A
. E P
E S Q

FIG 7

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATGGGCAGCA GCCATCATCA TCATCATCAC AGCAGCGGCC TGGTGCCGCG      50
M  G  S  S  H  H  H  H  H  H  S  S  G  L  V  P  R

CGGCAGCCAT ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCC     100
 G  S  H  M  A  S  M  T  G  G  Q  Q  M  G  R  I  L

TAGAACGTAT TCTGGAGAAT TGGGACCAAT GTGACACTCA GACGCTAAGA     150
  E  R  I  L  E  N  W  D  Q  C  D  T  Q  T  L  R

AAGAAACGAT TTATATTCTT CTGCAGTACC GCCTGGCCAC AATATCCTCT     200
K  K  R  F  I  F  F  C  S  T  A  W  P  Q  Y  P  L

TCAAGGGAGA GAAACCTGGC TTCCTGAGGG AAGTATAAAT TATAACATCA     250
 Q  G  R  E  T  W  L  P  E  G  S  I  N  Y  N  I  I

TCTTACAGCT AGACCTCTTC TGTAGAAAGG AGGGCAAATG GAGTGAAGTG     300
  L  Q  L  D  L  F  C  R  K  E  G  K  W  S  E  V

CCATATGTGC AAACTTTCTT TTCATTAAGA GACAACTCAC AATTATGTAA     350
P  Y  V  Q  T  F  F  S  L  R  D  N  S  Q  L  C  K

AAAGTGTGGT TTATGCCCTA CAGGAAGCCC TCAGAGTCCA CCTCCCTACC     400
 K  C  G  L  C  P  T  G  S  P  Q  S  P  P  P  Y  P

CCAGCGTCCC CTCCCCGACT CCTTCCTCAA CTAATAAGGA CCCCCCTTTA     450
  S  V  P  S  P  T  P  S  S  T  N  K  D  P  P  L

ACCCAAACGG TCCAAAAGGA GATAGACAAA GGGGTAAACA ATGAACCAAA     500
T  Q  T  V  Q  K  E  I  D  K  G  V  N  N  E  P  K

GAGTGCCAAT ATTCCCCGAT TATGCCCCCT CCAAGCAGTG AGAGGAGGAG     550
 S  A  N  I  P  R  L  C  P  L  Q  A  V  R  G  G  E

AATTCGGCCC AGCCAGAGTG CCTGTACCTT TTTCTCTCTC AGACTTAAAG     600
  F  G  P  A  R  V  P  V  P  F  S  L  S  D  L  K

CAAATTAAAA TAGACCTAGG TAAATTCTCA GATAACCCTG ACGGCTATAT     650
Q  I  K  I  D  L  G  K  F  S  D  N  P  D  G  Y  I

TGATGTTTTA CAAGGGTTAG GACAATCCTT TGATCTGACA TGGAGAGATA     700
 D  V  L  Q  G  L  G  Q  S  F  D  L  T  W  R  D  I

TAATGTTACT ACTAAATCAG ACACTAACCC CAAATGAGAG AAGTGCCGCT     750
  M  L  L  L  N  Q  T  L  T  P  N  E  R  S  A  A

GTAACTGCAG CCCGAGAGTT TGGCGATCTT TGGTATCTCA GTCAGGCCAA     800
V  T  A  A  R  E  F  G  D  L  W  Y  L  S  Q  A  N
```

FIG 7 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
CAATAGGATG ACAACAGAGG AAAGAACAAC TCCCACAGGC CAGCAGGCAG   850
 N R M  T T E E  R T T  P T G  Q Q A V

TTCCCAGTGT AGACCCTCAT TGGGACACAG AATCAGAACA TGGAGATTGG   900
  P S V  D P H  W D T E  S E H  G D W

TGCCACAAAC ATTTGCTAAC TTGCGTGCTA GAAGGACTGA GGAAAACTAG   950
C H K H  L L T  C V L  E G L R  K T R

GAAGAAGCCT ATGAATTACT CAATGATGTC CACTATAACA CAGGGAAAGG  1000
 K K P  M N Y S  M M S  T I T  Q G K E

AAGAAAATCT TACTGCTTTT CTGGACAGAC TAAGGGAGGC ATTGAGGAAG  1050
  E N L  T A F  L D R L  R E A  L R K

CATACCTCCC TGTCACCTGA CTCTATTGAA GGCCAACTAA TCTTAAAGGA  1100
H T S L  S P D  S I E  G Q L I  L K D

TAAGTTTATC ACTCAGTCAG CTGCAGACAT TAGAAAAAAC TTCAAAAGTC  1150
 K F I  T Q S A  A D I  R K N  F K S L

TGCCTAAGCT TGCGGCCGCA CTCGAGCACC ACCACCACCA CCACTGAGAT  1200
  P K L  A A A  L E H H  H H H  H . D

CCGGCTGCTA ACAAAGCCCG AAAGGAAGCT GAGTTGGCTN GTGGCNA    1247
P A A N  K A R  K E A  E L A X  G
```

FIG 8

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCC TAGAACGTAT    50
M A S M  T G G  Q Q M  G R I L  E R I

TCTGGAGAAT TGGGACCAAT GTGACACTCA GACGCTAAGA AAGAAACGAT   100
 L E N  W D Q C  D T Q  T L R  K K R F

TTATATTCTT CTGCAGTACC GCCTGGCCAC AATATCCTCT TCAAGGGAGA   150
  I F F  C S T  A W P Q  Y P L  Q G R

GAAACCTGGC TTCCTGAGGG AAGTATAAAT TATAACATCA TCTTACAGCT   200
E T W L  P E G  S I N  Y N I I  L Q L

AGACCTCTTC TGTAGAAAGG AGGGCAAATG GAGTGAAGTG CCATATGTGC   250
 D L F  C R K E  G K W  S E V  P Y V Q

AAACTTTCTT TTCATTAAGA GACAACTCAC AATTATGTAA AAAGTGTGGT   300
  T F F  S L R  D N S Q  L C K  K C G

TTATGCCCTA CAGGAAGCCC TCAGAGTCCA CCTCCCTACC CCAGCGTCCC   350
L C P T  G S P  Q S P  P P Y P  S V P

CTCCCCGACT CCTTCCTCAA CTAATAAGGA CCCCCCTTTA ACCCAAACGG   400
 S P T  P S S T  N K D  P P L  T Q T V

TCCAAAAGGA GATAGACAAA GGGGTAAACA ATGAACCAAA GAGTGCCAAT   450
  Q K E  I D K  G V N N  E P K  S A N

ATTCCCCGAT TATGCCCCCT CCAAGCAGTG AGAGGAGGAG AATTCGGCCC   500
I P R L  C P L  Q A V  R G G E  F G P

AGCCAGAGTG CCTGTACCTT TTTCTCTCTC AGACTTAAAG CAAATTAAAA   550
 A R V  P V P F  S L S  D L K  Q I K I

TAGACCTAGG TAAATTCTCA GATAACCCTG ACGGCTATAT TGATGTTTTA   600
  D L G  K F S  D N P D  G Y I  D V L

CAAGGGTTAG GACAATCCTT TGATCTGACA TGGAGAGATA TAATGTTACT   650
Q G L G  Q S F  D L T  W R D I  M L L

ACTAAATCAG ACACTAACCC CAAATGAGAG AAGTGCCGCT GTAACTGCAG   700
 L N Q  T L T P  N E R  S A A  V T A A

CCCGAGAGTT TGGCGATCTT TGGTATCTCA GTCAGGCCAA CAATAGGATG   750
  R E F  G D L  W Y L S  Q A N  N R M

ACAACAGAGG AAAGAACAAC TCCCACAGGC CAGCAGGCAG TTCCCAGTGT   800
T T E E  R T T  P T G  Q Q A V  P S V
```

FIG 8 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
AGACCCTCAT TGGGACACAG AATCAGAACA TGGAGATTGG TGCCACAAAC   850
 D  P  H   W  D  T  E   S  E  H   G  D  W   C  H  K  H

ATTTGCTAAC TTGCGTGCTA GAAGGACTGA GGAAAACTAG GAAGAAGCCT   900
  L  L  T   C  V  L   E  G  L  R   K  T  R   K  K  P

ATGAATTACT CAATGATGTC CACTATAACA CAGGGAAAGG AAGAAAATCT   950
 M  N  Y  S   M  M  S   T  I  T   Q  G  K  E   E  N  L

TACTGCTTTT CTGGACAGAC TAAGGGAGGC ATTGAGGAAG CATACCTCCC  1000
 T  A  F   L  D  R  L   R  E  A   L  R  K   H  T  S  L

TGTCACCTGA CTCTATTGAA GGCCAACTAA TCTTAAAGGA TAAGTTTATC  1050
  S  P  D   S  I  E   G  Q  L  I   L  K  D   K  F  I

ACTCAGTCAG CTGCAGACAT TAGAAAAAAC TTCAAAAGTC TGCCTAAGCT  1100
 T  Q  S  A   A  D  I   R  K  N   F  K  S  L   P  K  L

TGCGGCCGCA CTCGAGCACC ACCACCACCA CCACTGAGAT CCGGCTGCTA  1150
 A  A  A   L  E  H  H   H  H  H   H  .  D   P  A  A  N

ACAAAGCCCG AAAGGAAGCT GAGTTGGCTG GTGGCA                 1186
  K  A  R   K  E  A   E  L  A  G   G
```

FIG 9

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT      50
C P L C  S . S  S T G  A H C L  S Q L
 V R C  A P D P  A Q A  P I A  S P N W
  S A V  L L I  Q H R R  P L P  L P I

GGGCTAAAGG CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTCATC     100
 G . R  L A I V  P A Q  L S A  W V H P
  A K G  L P L  F L H S  . V P  G F I
G L K A  C H C  S C T  A K C L  G S S

CTAATCGAGC TGAACACTAG TCACTGGGTT CCACGGTTCT CTTCCATGAC     150
  N R A  E H .  S L G S  T V L  F H D
L I E L  N T S  H W V  P R F S  S M T
 . S S  . T L V  T G F  H G S  L P . P

CCATGGCTTC TAATAGAGCT ATAACACTCA CTGCATGGTC CAAGATTCCA     200
P W L L  I E L  . H S  L H G P  R F H
 H G F  . . S Y  N T H  C M V  Q D S I
  M A S  N R A  I T L T  A W S  K I P

TTCCTTGGAA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA ACACAAGGCT     250
 S L E  S V R P  R T P  G Q R  T Q G L
  P W N  P . D  Q E P Q  V R E  H K A
F L G I  R E T  K N P  R S E N  T R L

TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGGAAGC AGCCCGCCAC     300
  P P C  W K Q  P T T I  L E A  A R H
C H H V  G S S  P P P  F W K Q  P A T
 A T M  L E A A  H H H  F G S  S P P L

TATCTTGGGA GCTCTGGGAG CAAGGACCCC AGGTAACAAT TTGGTGACCA     350
Y L G S  S G S  K D P  R . Q F  G D H
 I L G  A L G A  R T P  G N N  L V T T
  S W E  L W E  Q G P Q  V T I  W . P

CGAAGGGACC TGAATCCGCA ACCATGAAGG GATCTCCAAA GCAATTGGAA     400
 E G T  . I R N  H E G  I S K  A I G N
  K G P  E S A  T M K G  S P K  Q L E
R R D L  N P Q  P . R  D L Q S  N W K
```

FIG 9 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATGTTCCTCC CAAGGCAAAA ATGCCCCTAA GATGTATTCT GGAGAATTGG    450
  V  P  P  K  A  K  M  P  L  R  C  I  L  E  N  W
M  F  L  P  R  Q  K  C  P  .  D  V  F  W  R  I  G
 C  S  S  Q  G  K  N  A  P  K  M  Y  S  G  E  L  G

GACCAATTTG ACCCTCAGAC AGTAAGAAAA AAATGACTTA TATTCTTCTG    500
D  Q  F  D  P  Q  T  V  R  K  K  .  L  I  F  F  C
 T  N  L  T  L  R  Q  .  E  K  N  D  L  Y  S  S  A
  P  I  .  P  S  D  S  K  K  K  M  T  Y  I  L  L

CAGTACCGCC CTGGCCACGA TATCCTCTTC AAGGGGGAGA AACCTGGCCT    550
 S  T  A  L  A  T  I  S  S  S  R  G  R  N  L  A  S
  V  P  P  W  P  R  Y  P  L  Q  G  G  E  T  W  P
Q  Y  R  P  G  H  D  I  L  F  K  G  E  K  P  G  L

CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACCTGTTTTG    600
  .  G  K  Y  K  L  .  H  H  L  T  A  R  P  V  L
P  E  G  S  I  N  Y  N  T  I  L  Q  L  D  L  F  C
 L  R  E  V  .  I  I  T  P  S  Y  S  .  T  C  F  V

TAGAAAAGGA GGCAAATGGA GTGAAGTGCC ATATTTACAA ACTTTCTTTT    650
.  K  R  R  Q  M  E  .  S  A  I  F  T  N  F  L  F
 R  K  G  G  K  W  S  E  V  P  Y  L  Q  T  F  F  S
  E  K  E  A  N  G  V  K  C  H  I  Y  K  L  S  F

CATTAAAAGA CAACTCGCAA TTATGTTAAC AGTGTGATTT GTGTTCCTAC    700
 I  K  R  Q  L  A  I  M  L  T  V  .  F  V  F  L  H
  L  K  D  N  S  Q  L  C  .  Q  C  D  L  C  S  Y
H  .  K  T  T  R  N  Y  V  N  S  V  I  C  V  P  T

ACGGAAGCCC TCAGATTCTA CTCCCCACCC CCGGCATCTC CCCTGAATCC    750
  G  S  P  Q  I  L  L  P  T  P  G  I  S  P  E  S
T  E  A  L  R  F  Y  S  P  P  P  A  S  P  L  N  P
 R  K  P  S  D  S  T  P  H  P  R  H  L  P  .  I  P

CTCCCCAACT TATT                                           764
L  P  N  L
 S  P  T  Y
  P  Q  L  I
```

FIG 10

```
      10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT      50
C P L C  S . S  S T G  A H C L  S Q L
 V R C  A P D P  A Q A  P I A  S P N W
  S A V  L L I  Q H R R  P L P  L P I

GGGCTAAAGG CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTCATC     100
 G . R  L A I V  P A Q  L S A  W V H P
  A K G  L P L  F L H S  . V P  G F I
G L K A  C H C  S C T  A K C L  G S S

CTAATCGAGC TGAACACTAG TCACTGGGTT CCACGGTTCT CTTCCATGAC     150
  N R A  E H .  S L G S  T V L  F H D
L I E L  N T S  H W V  P R F S  S M T
 . S S  . T L V  T G F  H G S  L P . P

CCATGGCTTC TAATAGAGCT ATAACACTCA CTGCATGGTC CAAGATTCCA     200
P W L L  I E L  . H S  L H G P  R F H
 H G F  . . S Y  N T H  C M V  Q D S I
  M A S  N R A  I T L T  A W S  K I P

TTCCTTGGAA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA ACACAAGGCT     250
 S L E  S V R P  R T P  G Q R  T Q G L
  P W N  P . D  Q E P Q  V R E  H K A
F L G I  R E T  K N P  R S E N  T R L

TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGGAAGC GGCCCGCCAC     300
  P P C  W K Q  P T T I  L E A  A R H
C H H V  G S S  P P P  F W K R  P A T
 A T M  L E A A  H H H  F G S  G P P L

TATCTTGGGA GCTCTGGGAG CAAGGACCCC CAGGTAACAA TTTGGTGACC     350
Y L G S  S G S  K D P  Q V T I  W . P
 I L G  A L G A  R T P  R . Q  F G D H
  S W E  L W E  Q G P P  G N N  L V T

ACGAAGGGAC CTGAATCCGC AACCATGAAG GGATCTCCAA AGCAATTGGA     400
 R R D  L N P Q  P . R  D L Q  S N W K
  E G T  . I R  N H E G  I S K  A I G
T K G P  E S A  T M K  G S P K  Q L E
```

FIG 10 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
AATGTTCCTC CCAAGGCAAA AATGCCCCTA AGATGTATTC TGGAGAATTG   450
  C S S  Q G K  N A P K  M Y S  G E L
N V P P  K A K  M P L  R C I L  E N W
 M F L  P R Q K  C P .  D V F  W R I G

GGACCAATCT GACCCTCAGA CAGTAAGAAA AAAAATGACT TATATTCTTC   500
G P I .  P S D  S K K  K N D L  Y S S
 D Q S  D P Q T  V R K  K M T  Y I L L
  T N L  T L R  Q . E K  K . L  I F F

TGCAGTACCG CCTGGCCACG GATATCCTCT TCAAGGGGGA GAAACCTGGC   550
 A V P  P G H G  Y P L  Q G G  E T W P
  Q Y R  L A T  D I L F  K G E  K P G
C S T A  W P R  I S S  S R G R  N L A

CTCCTGAGGG AAGTATAAAT TATAACACCA TCTTACAGCT AGACCTGTTT   600
  P E G  S I N  Y N T I  L Q L  D L F
L L R E  V . I  I T P  S Y S .  T C F
 S . G  K Y K L  . H H  L T A  R P V L

TGTAGAAAAG GAGGCAAATG GAGTGAAGTG CCATATTTAC AAACTTTCTT   650
C R K G  G K W  S E V  P Y L Q  T F F
 V E K  E A N G  V K C  H I Y  K L S F
  . K R  R Q M  E . S  A I F T  N F L

TTCATTAAAA GACAACTCGC AATTATGTAA ACAGTGTGAT TTGTGTCCTA   700
 S L K  D N S Q  L C K  Q C D  L C P T
  H . K  T T R  N Y V N  S V I  C V L
F I K R  Q L A  I M .  T V .  F V S Y

CAGGAAGCCC TCAGATCTAC CTCCCTACCC CGGCATCTCC CTGACTCCTT   750
  G S P  Q I Y  L P T P  A S P  . L L
Q E A L  R S T  S L P  R H L P  D S F
 R K P  S D L P  P Y P  G I S  L T P S

CCCCAACTAA TAAGGACCCA CTTCAGCCCA AACAGTCCAA AAGGACATAG   800
P Q L I  R T H  F S P  N S P K  G H
 P N .  . G P T  S A Q  T V Q  K D I
  P T N  K D P  L Q P K  Q S K  R T .
```

FIG 11

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
GGCATTGATA GCACCCATCA GATGGCCAAA TCATTATTTA CTGGACCAGG      50
G I D S  T H Q  M A K  S L F T  G P G
 A L I  A P I R  W P N  H Y L  L D Q A
  H . .  H P S  D G Q I  I I Y  W T R

CCTTTTCAAA ACTATCAAGC AGATAGGGCC CGTGAAGCAT GCCAAAGAAA     100
 L F K  T I K Q  I G P  V K H  A K E I
  F S K  L S S  R . G P  . S M  P K K
P F Q N  Y Q A  D R A  R E A C  Q R N

TAATCCCCTG CCTTATCGCC ATGTTCCTTC AGGAGAACAA AGAACAGGCC     150
  I P C  L I A  M F L Q  E N K  E Q A
. S P A  L S P  C S F  R R T K  N R P
 N P L  P Y R H  V P S  G E Q  R T G H

ATTACCCAGG GGAAGACTGG CAACTAGATT TTACCCACAT GGCCAAATGT     200
I T Q G  K T G  N . I  L P T W  P N V
 L P R  G R L A  T R F  Y P H  G Q M S
  Y P G  E D W  Q L D F  T H M  A K C

CAGGGATTTC AGCATCTACT AGTCTGGGCA GATACTTTCA CTGGTTGGGT     250
 R D F  S I Y .  S G Q  I L S  L V G W
  G I S  A S T  S L G R  Y F H  W L G
Q G F Q  H L L  V W A  D T F T  G W V

GGAGTCTTCT CCTTGTAGGA CAGAAAAGAC CCAAGAGGTA ATAAAGGCAC     300
  S L L  L V G  Q K R P  K R .  . R H
G V F S  L . D  R K D  P R G N  K G T
 E S S  P C R T  E K T  Q E V  I K A L

TAATGAAATA ATTCCCAGAT TTGGACTTCC CCCAGGATTA CAGGGTGACA     350
. . N N  S Q I  W T S  P R I T  G . Q
 N E I  I P R F  G L P  P G L  Q G D N
  M K .  F P D  L D F P  Q D Y  R V T
```

FIG 11 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATGGCCCCGC TTTCAAGGCT GCAGTAACCC AGGGAGTATC CCAGGTGTTA    400
 W  P  R  F  Q  G  C  S  N  P  G  S  I  P  G  V  R
  G  P  A  F  K  A  A  V  T  Q  G  V  S  Q  V  L
M  A  P  L  S  R  L  Q  .  P  R  E  Y  P  R  C  .

GGCATACAAT ATCACTTACA CTGTGCCTGG AGGCCACAAT CCTCCAGAAA    450
  H  T  I  S  L  T  L  C  L  E  A  T  I  L  Q  K
G  I  Q  Y  H  L  H  C  A  W  R  P  Q  S  S  R  K
 A  Y  N  I  T  Y  T  V  P  G  G  H  N  P  P  E  K

AGTCAAGAAA ATGAATGAAA CACTCAAAGA TCTAAAAAAG CTAACCCAAG    500
S  Q  E  N  E  .  N  T  Q  R  S  K  K  A  N  P  R
 V  K  K  M  N  E  T  L  K  D  L  K  K  L  T  Q  E
  S  R  K  .  M  K  H  S  K  I  .  K  S  .  P  K

AAACCCACAT TGCATGACCT GTTCTGTTGC CTATAACCTT ACTAAGAATC    550
 N  P  H  C  M  T  C  S  V  A  Y  N  L  T  K  N  P
  T  H  I  A  .  P  V  L  L  P  I  T  L  L  R  I
K  P  T  L  H  D  L  F  C  C  L  .  P  Y  .  E  S

CATAACTATC CCCCAAAAAG CAGGACTTAG CCCATACGAG ATGCTATATG    600
  .  L  S  P  K  K  Q  D  L  A  H  T  R  C  Y  M
H  N  Y  P  P  K  S  R  T  .  P  I  R  D  A  I  W
 I  T  I  P  Q  K  A  G  L  S  P  Y  E  M  L  Y  G

GATGGCCTTT CCTAACCAAT GACCTTGTGC TTGACTGAGA AATGGCCAAC    650
D  G  L  S  .  P  M  T  L  C  L  T  E  K  W  P  T
 M  A  F  P  N  Q  .  P  C  A  .  L  R  N  G  Q  L
  W  P  F  L  T  N  D  L  V  L  D  .  E  M  A  N

TTAGTTGCAG ACATCACCTC CTTAGCCAAA TATCAACAAG TTCTTAAAAC    700
 .  L  Q  T  S  P  P  .  P  N  I  N  K  F  L  K  H
  S  C  R  H  H  L  L  S  Q  I  S  T  S  S  .  N
L  V  A  D  I  T  S  L  A  K  Y  Q  Q  V  L  K  T
```

FIG 11 (suite)

```
        10         20         30         40         50
1234567890 1234567890 1234567890 1234567890 1234567890
ATCACAGGGA ACCTGTCCCC GAGAGGAGGG AAAGGAACTA TTCCACCCTG     750
  H R E   P V P   E R R E   R N Y   S T L
I T G N   L S P   R G G   K G T I   P P W
 S Q G   T C P R   E E G   K E L   F H P G

GTGACATG                                                  758
V T
 . H
  D M
```